IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of: Ruben et al.

Application Serial No.: 09/880,748 — Group Art Unit: 1631

Filed: June 15, 2001 — Examiner: DUFFY, Patricia

Title: ANTIBODIES THAT IMMUNOSPECIFICALLY BIND TO B LYMPHOCYTE STIMULATOR PROTEIN (as amended) — Atty. Docket No. PF523P1

DECLARATION OF RODGER SMITH UNDER 37 C.F.R. § 1.132

Commissioner for Patents
Alexandria, VA 22313-1450

Sir:

I, Rodger Smith Ph.D., hereby declare and state as follows:

1. I am currently employed as a Senior Scientist I in Lead Identification at CoGenesys, a division of Human Genome Sciences, Inc. I understand that Human Genome Sciences is the assignee of the above-captioned patent application. I earned my Ph.D. in 1989 from the Department of Microbiology at the University of Illinois, Urbana-Champaign, Illinois. It was during my thesis research that I first began work on cloning and sequencing of antibody genes. From 1990 to 1999, I worked as a Scientist in the Molecular Biology and Assay Development groups at IGEN International, Inc. (now known as Bioveris Corp.) where my primary responsibilities were developing and characterizing antibody reagents for therapeutic and diagnostic applications. A portion of this work entailed the design and construction of both human and mouse V-domain antibody repertoire libraries for display on the surface of bacteriophage including in 1997 an SBIR grant sponsored by The Department of the Army to construct and validate a large semi-synthetic human phage antibody display library. In 2000, I joined the Antibody Development group at Human Genome Sciences where I have continued to work with phage antibody display technology, primarily for developing therapeutic antibodies to a

variety of novel protein targets. A large portion of this work involved screening and characterizing hundreds of antibody leads at both the DNA sequence and protein level. I am the co-author of 12 scientific articles and several issued and pending patent applications. A copy of my curriculum vitae is attached as Exhibit H[1].

2. On December 14, 2004, I signed a Declaration (hereafter "the December 2004 Declaration") explaining that an antibody scientist on or before June 16, 2000 would have been able to identify and correct errors present in Table 1 of the 09/880,748 application (hereafter "the '748 Application"). The errors pertained to the delineation of the VL region of the amino acid sequence of certain scFvs disclosed in SEQ ID NOS:1-2128. I understand that the December 2004 Declaration was submitted to the United States Patent and Trademark Office (hereafter "Patent Office") on December 14, 2004 to support the permissibility of correcting the errors in delineating the VL region in column 3 of Table 1 of the '748 application. I understand the Patent Office has not allowed the corrections to Table 1 because the Patent Office was not convinced that an antibody scientist would have been able to both recognize and correct the errors in Table 1 of the '748 application as of June 16, 2000.

3. I have read and examined the Examiner's comments relating to the December 2004 Declaration as set forth in the communication from the Patent Office mailed May 5, 2005 on pages 5-11 under the heading "New Objections or Rejections based on Amendment" (hereafter "the comments mailed May 5, 2005"). I maintain that an antibody scientist, indeed, would have been able to both recognize and correct the errors in the delineation of the VL region in column 3 of Table 1 of the '748 application as of June 16, 2000. Further, to convince the Patent Office of the accuracy of my position, I believe that certain points from the December 2004 Declaration warrant clarification.

---

[1] On December 14, 2004, I signed a Declaration under 37 C.F.R. 1.132 ("the December 2004 Declaration") which was submitted to the United States Patent and Trademark Office in connection with this case. The December 2004 Declaration was submitted with 7 Exhibits labeled Exhibits A-G. Copies of Exhibits that were previously submitted in conjunction with the December 2004 Declaration, and that are referred to in the present Declaration are attached hereto using the same Exhibit Letter as in the December 2004 Declaration. Exhibits newly submitted in conjunction with this Declaration are consecutively lettered from H-M. Although a copy of my curriculum vitae (CV) was submitted with the December 2004 Declaration as Exhibit A, I have revised my CV to reflect my transfer within HGS to the CoGenesys division since my signing of the December 2004 Declaration. My updated CV is submitted herewith as Exhibit H.

SUMMARY OF PREVIOUS DECLARATION

4. In the December 2004 Declaration, I stated that an antibody scientist would both be able to identify and correct errors in the delineation of the VL domain in Table 1 of the 09/880,748.. The premise of the December 2004 Declaration was that an antibody scientist can (and could have as of June 16, 2000) recognize and correct the errors in Table 1 on the basis of an art accepted standard numbering system for the variable regions of kappa and lambda light chains. In paragraph 9 of the December 2004 Declaration, I stated:

> The beginning of the VL region in an scFv may be easily delineated by 1) determining whether the scFv contains a kappa or a lambda variable domain and then 2) calculating the first amino acid sequence based on a standard numbering system for immunoglobulin variable regions that was established by Elvin A. Kabat and Tai Te Wu in the 1970's that is widely used by immunologists even today.

5. In the comments mailed May 5, 2005, concerns were raised by the Patent Office regarding both steps, i.e., the first step of recognizing if a light chain variable domain is a kappa variable domain or a lambda variable domain and the second step of applying the Kabat-Wu numbering system. Below, I make clarifications on both issues.

Determining if the VL in the scFv is a Vκ or a Vλ

6. Paragraphs 10-13 of the December 2004 Declaration explained that an antibody scientist could routinely determine if a VL domain was a Vκ or a Vλ. In paragraph 13, I stated that one of skill in the art could accomplish this task by aligning the sequence of an scFv against a database containing known human germline genes and identifying whether the "closest germline" gene is a Vκ or a Vλ.

7. To clarify, the purpose of aligning a given VL domain to known germline genes is not to "identify the closest germline gene" *per se*. Rather, the point is to determine if the VL domain in question is most similar to Vκ or Vλ variable regions because if a VL domain in question is most similar to Vκ variable regions, it too is a Vκ, whereas if a the VL domain in question is most similar to Vλ variable regions, it is a Vλ.

The ability to classify a VL gene as a either a Vκ or Vλ via alignment to known VL sequences is based on the principle that Vκs are more similar to other Vκs than they are to Vλs and vice versa. This principle is set forth in the 1993 textbook, *Fundamental Immunology*:

> Variable region sequences do not randomly differ relative to each other – even within the CDRs. **Analyses of hundreds of V regions reveal that sequences naturally fall into a homology based hierarchy directly related to the germline antibody gene loci. Members within a hierarchical group are more similar to each other than to all sequences from other groups**; furthermore, similar sequences display a shared pattern of amino acid substitutions that serve as "membership badges" for the various classifications. Through examination of these linked-substitutions, an evolutionary history of V regions can be obtained. Of course, the oldest and most basic group is that of the V regions themselves, followed by the division into VH, Vκ and Vλ representing separate V-gene loci on different chromosomes. (emphasis added, *Fundamental Immunology,* 3rd edition, edited by William Paul, Raven Press, New York. 1993: p290, Exhibit I)

8. Therefore, in view of the overarching principle that Vκs are more similar to other Vκs than they are to Vλs and vice versa, an antibody scientist can, and could have used as June 16, 2000, use such routine alignments to classify a VL domain as either a Vκ or Vλ. For this purpose, any database containing human light chain germline variable region genes could be used, not just the database listed in the paragraph [0669] of the '748 application. Furthermore, for this purpose, it is not necessary to definitively establish the amino acid sequence of the actual germline gene.

Kabat-Wu Numbering System

9. Paragraphs 14-15 of the December 2004 Declaration explained that an art accepted numbering system for immunological variable domains developed by Elvin A. Kabat and Tai Te Wu in the 1970s could be used to identify the first amino acid residue in a VL domain.

10. In brief, the Kabat-Wu numbering system recognizes the presence of an invariant cysteine residue at position 23 of immunoglobulin light chain variable regions.

The Kabat-Wu numbering system also acknowledges that lambda light chains have a deletion at Kabat-Wu position 10, so the invariant cysteine residue at Kabat-Wu's position 23 in lambda light chains is actually the 22nd amino acid residue in the Vλ domain. At paragraph 15 of the December 2004 Declaration, I stated that on the basis of this information, an antibody scientist could identify amino acid residue number 1 in a VL domain by identifying the invariant cysteine residue at Kabat-Wu position 23, assigning that cysteine as amino acid residue number 23 or 22 depending on whether the VL domain was a Vκ or Vλ, respectively, and counting backwards to amino acid residue 1 to identify the first amino acid residue of the VL region.

11. In the December 14, 2004 Declaration, I supplied a copy of Table 1 from the Introduction of the fifth edition of *Sequences of Proteins of Immunological Interest* as documentation of the Kabat-Wu numbering system (Exhibit G). Table 1 from *Sequences of Proteins of Immunological Interest* indicates that there is an "occasional" amino acid residue at position number 0 in immunoglobulin light chains. In the comments mailed May 5, 2005, the Patent Office took the position that the amino acid residues that were being excluded from the delineation of the VL region in Table 1 of the '748 application by virtue of the requested corrections could be construed to be examples of VL regions which had an amino acid residue at Kabat-Wu position 0.

12. I disagree with this position taken by the Patent Office. I do not recall ever working with a human light chain variable region that contained an amino acid residue at position 0, nor would I expect to work with one because human germline variable region sequences do not contain an amino acid at position number 0 (e.g., see Exhibits C and D). Based on my experience, I would characterize the human light chain variable region with an amino acid residue at position number 0 as "extremely rare."

13. Unfortunately, the *Sequences of Proteins of Immunological Interest* does not further define the meaning of the "occasional" amino acid residue at position 0 in terms of a frequency of occurrence. Nonetheless, given the sheer volume of known immunoglobulin VL region sequences, the frequency of human VL domains with an amino acid residue at position 0 can be better defined empirically as set forth below.

Frequency of human VL domains with an amino acid residue at position 0

14. In addition to establishing a standard numbering system for immunoglobulin variable region sequences, Kabat-Wu also established a database that contains thousands of examples of immunoglobulin sequences (See Exhibit F for a brief description of this database).

15. For the exercise described below, an online database, known as the KabatMan database was used. This database is available at the website of Dr. Andrew Martin, a bioinformaticist and professor at the University College of London. The URL for this website is: http://www.bioinf.org.uk/abs/simkab.html. The KabatMan database contains sequences from the July 12, 2000 version of the Kabat-Wu database. The contents of the KabatMan database are described on the website as follows, "Only immunoglobulin sequences are stored here; all T-cell receptor, MHC sequences, etc. are rejected. In addition, sequences with fewer than 75 residues are rejected so that the database contains only essentially complete light or heavy chain sequences."

16. The KabatMan database was queried to provide a listing of all the human light chain sequences it contained. Prior to analysis, any sequences that were N-terminally truncated were removed from the dataset, as were any redundant amino acid sequences. The dataset analyzed contained 1,745 immunoglobulin light chain sequences. Of these, only 5/1745 or 0.29% contained an amino acid residue at position 0. These sequences are shown in Exhibit J.

17. This 0.29% frequency was so low as to suggest that the sequences containing an amino acid residue at position number 0 were in error. Both the Kabat-Wu database and the KabatMan derivative of the Kabat-Wu database provide information as to the original source of each sequence contained in the databases. Moreover, it is suggested by the current curators of the Kabat database that "if there are doubts about these sequences or their annotations, please refer to the original papers" (Exhibit F, top of the right hand column on page 214). Because the frequency of the position 0 amino acid residue was so low in human VL domains as to suggest error, the original papers presenting these sequences were consulted. Exhibit J also lists the original citations associated with each sequence.

18. The 5 sequences identified in paragraph 16 above as containing a position 0 amino acid residue are the HBL-2' CL, HBL-3' CL, MP9' CL, RPMI8226'CL and the MC116'CL sequences. These 5 sequences were originally published in three scientific articles. In each case, the presence of the amino acid residue at position 0 in the database sequence turned out to be an error.

The HBL-2' CL and the HBL-3' CL sequences

19. The HBL-2' CL and the HBL-3' CL sequences were originally presented in a 1994 article published in *Blood* by Riboldi et al. (submitted herewith as K). The alanine (A) amino acid residue at position 0 shown in the sequence in the KabatMan database, is simply not present in the original sequence (see Figure 2B). Nor is the position 0 alanine for these sequences present in the GenBank Reports for these sequences[2]. Thus, it is clear that these two examples of position 0 amino acids are simply inaccurate.

The MP9'CL sequence

20. The MP9'CL sequence was originally presented in a 1995 article published in *Molecular Immunology* by Andris et al. (submitted herewith as Exhibit L). The aspartic acid (D) amino acid residue at position 0 shown in the sequence in the KabatMan database is present in the sequence in the original paper (see Figure 3C), but careful reading of the paper shows that this amino acid residue is actually encoded by the vector sequence into which the light chain was cloned prior to sequencing. Figure 3C indicates that the MP9 variable region is a VΛ2 variable region. Table 2 indicates that the PCR primer used to amplify this sequence would have been the VLAM 2 primer and that this primer amplifies V lambda gene segments "from the *beginning* of framework 1." Reproduced below is the beginning of the MP9 sequence in which the VLAM2 primer sequence is underlined:

GAT TCG TAT CAG CTG ACG CAG CCT CCC TCC...
E S Y Q L T Q P P S...

The GAT codon encoding the position 0 amino acid codon is 5' of the PCR primer indicating that the source of this codon cannot be the light chain. Reading of the Materials and Methods section shows that the PCR products encoding the light chains were cloned

as blunt-ended fragments into the EcoRV restriction site of the Bluescript phagemid vector (See page 1108, section entitled "Isolation, cloning and sequencing of the amplified products", first sentence.) The EcoRV site is reproduced below with the cut site indicated by the forward slash mark.

GAT/ATC
CTA/TAG

Thus, the position 0 aspartic acid residue in the MP9'CL light chain in the KabatMan database is the result of the erroneous inclusion of the 5' portion of the vector-derived EcoRV restriction site.

The RPMI8226'CL and the MC116'CL sequences

21. The RPMI8226'CL and the MC116'CL sequences were originally presented in a 1995 article published in *The Scandinavian Journal of Immunology* by Watkins et al. (submitted herewith as Exhibit M). The glutamic acid amino residue acid at position 0 shown in these sequences in the Kabat database is present in the sequence in the original paper (see Table 2), but careful reading of the paper shows that this amino acid residue is actually encoded by the PCR primer used to amplify the light chain from the RPMI8226 and MC116 cell lines. It is reported at page 446, top of the left hand column, that the light chains expressed by the MC116 and RPMI8226 cell lines are both Vλ2 light chains, but that these sequences were best amplified with a primer meant to amplify Vλ1 light chains, primer 3a shown in Table 1. The legend to Table 1 indicates that the underlined sequence in each primer is a restriction site that was *inserted* into the PCR primer for cloning purposes. Reproduced below is the beginning of the RPMI8226'CL and MC116'CL sequences as shown in Table 2 of the Watkins paper in which the 3a primer sequence is underlined and the inserted restriction site is shown in boldfaced text.:

GAG	**CTC**	TCT	GTG	CTG	ACT	CAG	CCT	GCC...
E	L	S	V	L	T	Q	P	A...

Because the 5' end of each of these light chain sequences is the region to which the PCR primer annealed, the 5' end of this sequence is defined by the PCR primer and not the light chains found in the MC116 and RPMI8226 cell lines. Moreover, the glutamic acid amino

[2] Accession Numbers L29113 and L29114 are listed in the legend of Figure 2 of the Riboldi et al. reference as the GenBank Accession Numbers for the HBL-2' CL and the HBL-3' CL sequences.

acid residue at position 0 is encoded by the restriction site that was inserted into the PCR primer. Thus, the presence of the glutamic acid amino acid residue at position 0 in the RPMI8226'CL and the MC116'CL light chain sequences in the KabatMan database is an error.

22. Accordingly, 0/1745 human light chain variable region sequences in the KabatMan database contain a verifiable position 0 amino acid residue. Thus, the presence of the "occasional" position 0 amino acid in human VL regions is, as my previous experience taught me, either extremely rare or non-existent.

Correction of the Delineation of VL regions

23. I maintain that an antibody scientist as of June 16, 2000 would have been able to recognize and correct the errors in Table 1 of the '748 application relating to the delineation of the VL region in the scFvs of SEQ ID NOS:1-2128 on the basis of the Kabat–Wu numbering system. This is true whether the error was the inclusion of a few extra amino acids (usually A, AL of AF) or the omission of a serine (S) residue at the N terminal end of certain VL regions.

24. With particular regard to the omitted serine residues, an antibody scientist would certainly have made this correction even though the omitted serine residue to be included is also the last residue of the $(Gly_4Ser)_3$ linker sequence. An antibody scientist would make this correction on the basis of the Kabat–Wu numbering system established after careful analysis and alignment of numerous V region sequences and on the basis that it is accepted that human germline light chain variable regions have an amino acid residue at Kabat-Wu position 1[3].

[3] The information provided in paragraph 18 of the December 2004 Declaration describing how in 238/239 instances where a correction to delineate the VL region as including the last serine of the $(Gly_4Ser)_3$ linker sequence was requested, the closest identified germline gene was a V gene whose first amino acid residue was a serine was tangential information, meant by way of explanation and as a secondary proof that one would wish to include the residue. However, even in the case where the initial amino acid of the closest germline gene was not serine (*i.e.*, SEQ ID NO:1389), the correction would still be made by the antibody scientist as of June 16, 2000 to include the serine residue because of the mandates of the Kabat-Wu numbering system.

SUMMARY

25. On or before June 16, 2000, an antibody scientist examining the information presented in Table 1 and the sequences of SEQ ID NOS:1-2128 of the Sequence Listing of the '748 application would have readily recognized that in several instances, the amino acid residues delineated in Table 1 as making up the VL region of certain scFvs were incorrect for either containing a few additional amino acids at the amino terminal end of the VL region or for lacking an amino acid at the amino terminal end of the VL-region. Moreover, on the basis of the Kabat-Wu numbering system, an antibody scientist would also have been able to correct the delineations of the VL regions in Table 1 the '748 application. The corrections an antibody scientist would have made, are the same as those that were requested in the '748 application on December 14, 2000.

26. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under § 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application captioned above or any patent issuing thereupon.

Date: Aug. 4th, 2005

Rodger Smith

Rodger Smith, Ph.D.

Exhibit C

Although the human Ig kappa locus has not been fully sequenced yet, all of individual genes are likely to have been isolated (Schable KF and Zachau HG, 1993; Brensing-Kuppers J. et al, 1997). The following sequences are taken from these studies.

Total number of sequences: 46

>A1
DVVMTQSPLSLPVTLGQPASISCRSSQSLVYSDGNTYLNWFQQRPGQSPRRLIYKVSNWD
SGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGTHWP

>A10
EIVLTQSPDFQSVTPKEKVTITCRASQSIGSSLHWYQQKPDQSPKLLIKYASQSFSGVPS
RFSGSGSGTDFTLTINSLEAEDAATYYCHQSSSLP

>A11
EIVLTQSPATLSLSPGERATLSCGASQSVSSSYLAWYQQKPGLAPRLLIYDASSRATGIP
DRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSSP

>A14
DVVMTQSPAFLSVTPGEKVTITCQASEGIGNYLYWYQQKPDQAPKLLIKYASQSISGVPS
RFSGSGSGTDFTFTISSLEAEDAATYYCQQGNKHP

>A17
DVVMTQSPLSLPVTLGQPASISCRSSQSLVYSDGNTYLNWFQQRPGQSPRRLIYKVSNRD
SGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGTHWP

>A18
DIVMTQTPLSLSVTPGQPASISCKSSQSLLHSDGKTYLYWYLQKPGQSPQLLIYEVSSRF
SGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGIHLP

>A19
DIVMTQSPLSLPVTPGEPASISCRSSQSLLHSNGYNYLDWYLQKPGQSPQLLIYLGSNRA
SGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQALQTP

>A2
DIVMTQTPLSLSVTPGQPASISCKSSQSLLHSDGKTYLYWYLQKPGQPPQLLIYEVSNRF
SGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQSIQLP

>A20
DIQMTQSPSSLSASVGDRVTITCRASQGISNYLAWYQQKPGKVPKLLIYAASTLQSGVPS
RFSGSGSGTDFTLTISSLQPEDVATYYCQKYNSAP

>A23
DIVMTQTPLSSPVTLGQPASISCRSSQSLVHSDGNTYLSWLQQRPGQPPRLLIYKISNRF
SGVPDRFSGSGAGTDFTLKISRVEAEDVGVYYCMQATQFP

>A26
EIVLTQSPDFQSVTPKEKVTITCRASQSIGSSLHWYQQKPDQSPKLLIKYASQSFSGVPS
RFSGSGSGTDFTLTINSLEAEDAATYYCHQSSSLP

>A27
EIVLTQSPGTLSLSPGERATLSCRASQSVSSSYLAWYQQKPGQAPRLLIYGASSRATGIP
DRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSSP

>A3
DIVMTQSPLSLPVTPGEPASISCRSSQSLLHSNGYNYLDWYLQKPGQSPQLLIYLGSNRA
SGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQALQTP

>A30
DIQMTQSPSSLSASVGDRVTITCRASQGIRNDLGWYQQKPGKAPKRLIYAASSLQSGVPS
RFSGSGSGTEFTLTISSLQPEDFATYYCLQHNSYP

>A5
EIVMTQTPLSLSITPGEQASISCRSSQSLLHSDGYTYLYWFLQKARPVSTLLIYEVSNRF
SGVPDRFSGSGSGTDFTLKISRVEAEDFGVYYCMQDAQDPP

>A7
DIVMTQTPLSSPVTLGQPASISFRSSQSLVHSDGNTYLSWLQQRPGQPPRLLIYKVSNRF
SGVPDRFSGSGAGTDFTLKISRVEAEDVGVYYCTQATQFP

>B2
ETTLTQSPAFMSATPGDKVNISCKASQDIDDDMNWYQQKPGEAAIFIIQEATTLVPGIPP
RFSGSGYGTDFTLTINNIESEDAAYYFCLQHDNFP

>B3
DIVMTQSPDSLAVSLGERATINCKSSQSVLYSSNNKNYLAWYQQKPGQPPKLLIYWASTR
ESGVPDRFSGSGSGTDFTLTISSLQAEDVAVYYCQQYYSTP

>L1
DIQMTQSPSSLSASVGDRVTITCRASQGISNYLAWFQQKPGKAPKSLIYAASSLQSGVPS
RFSGSGSGTDFTLTISSLQPEDFATYYCQQYNSYP

>L10
EIVMTQSPPTLSLSPGERVTLSCRASQSVSSSYLTWYQQKPGQAPRLLIYGASTRATSIP
ARFSGSGSGTDFTLTISSLQPEDFAVYYCQQDHNLPP

>L11
AIQMTQSPSSLSASVGDRVTITCRASQGIRNDLGWYQQKPGKAPKLLIYAASSLQSGVPS
RFSGSGSGTDFTLTISSLQPEDFATYYCLQDYNYP

>L12
DIQMTQSPSTLSASVGDRVTITCRASQSISSWLAWYQQKPGKAPKLLIYDASSLESGVPS
RFSGSGSGTEFTLTISSLQPDDFATYYCQQYNSYS

>L14
NIQMTQSPSAMSASVGDRVTITCRARQGISNYLAWFQQKPGKVPKHLIYAASSLQSGVPS
RFSGSGSGTEFTLTISSLQPEDFATYYCLQHNSYP

>L15
DIQMTQSPSSLSASVGDRVTITCRASQGISSWLAWYQQKPEKAPKSLIYAASSLQSGVPS
RFSGSGSGTDFTLTISSLQPEDFATYYCQQYNSYP

>L16
EIVMTQSPATLSVSPGERATLSCRASQSVSSNLAWYQQKPGQAPRLLIYGASTRATGIPA
RFSGSGSGTEFTLTISSLQSEDFAVYYCQQYNNWP

>L18
AIQLTQSPSSLSASVGDRVTITCRASQGISSALAWYQQKPGKAPKLLIYDASSLESGVPS
RFSGSGSGTDFTLTISSLQPEDFATYYCQQFNNYP

>L19
DIQMTQSPSSVSASVGDRVTITCRASQGISSWLAWYQQKPGKAPKLLIYAASSLQSGVPS
RFSGSGSGTDFTLTISSLQPEDFATYYCQQANSFP

>L2
EIVMTQSPATLSVSPGERATLSCRASQSVSSNLAWYQQKPGQAPRLLIYGASTRATGIPA
RFSGSGSGTEFTLTISSLQSEDFAVYYCQQYNNWP

>L20
EIVLTQSPATLSLSPGERATLSCRASQGVSSYLAWYQQKPGQAPRLLIYDASNRATGIPA
RFSGSGPGTDFTLTISSLEPEDFAVYYCQQRSNWH

>L22
DIQMIQSPSFLSASVGDRVSIICWASEGISSNLAWYLQKPGKSPKLFLYDAKDLHPGVSS
RFSGRGSGTDFTLTIISLKPEDFAAYYCKQDFSYPP

>L23
AIRMTQSPFSLSASVGDRVTITCWASQGISSYLAWYQQKPAKAPKLFIYYASSLQSGVPS
RFSGSGSGTDYTLTISSLQPEDFATYYCQQYYSTP

>L24
VIWMTQSPSLLSASTGDRVTISCRMSQGISSYLAWYQQKPGKAPELLIYAASTLQSGVPS
RFSGSGSGTDFTLTISCLQSEDFATYYCQQYYSFP

>L25
EIVMTQSPATLSLSPGERATLSCRASQSVSSSYLSWYQQKPGQAPRLLIYGASTRATGIP
ARFSGSGSGTDFTLTISSLQPEDFAVYYCQQDYNLP

>L4/18a
AIQLTQSPSSLSASVGDRVTITCRASQGISSALAWYQQKPGKAPKLLIYDASSLESGVPS
RFSGSGSGTDFTLTISSLQPEDFATYYCQQFNSYP

>L5
DIQMTQSPSSVSASVGDRVTITCRASQGISSWLAWYQQKPGKAPKLLIYAASSLQSGVPS
RFSGSGSGTDFTLTISSLQPEDFATYYCQQANSFP

>L6
EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPA
RFSGSGSGTDFTLTISSLEPEDFAVYYCQQRSNWP

>L8
DIQLTQSPSFLSASVGDRVTITCRASQGISSYLAWYQQKPGKAPKLLIYAASTLQSGVPS
RFSGSGSGTEFTLTISSLQPEDFATYYCQQLNSYP

>L9
AIRMTQSPSSFSASTGDRVTITCRASQGISSYLAWYQQKPGKAPKLLIYAASTLQSGVPS
RFSGSGSGTDFTLTISCLQSEDFATYYCQQYYSYP

>O1
DIVMTQTPLSLPVTPGEPASISCRSSQSLLDSDDGNTYLDWYLQKPGQSPQLLIYTLSYR
ASGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQRIEFP

>O11
DIVMTQTPLSLPVTPGEPASISCRSSQSLLDSDDGNTYLDWYLQKPGQSPQLLIYTLSYR
ASGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQRIEFP

>O12
DIQMTQSPSSLSASVGDRVTITCRASQSISSYLNWYQQKPGKAPKLLIYAASSLQSGVPS
RFSGSGSGTDFTLTISSLQPEDFATYYCQQSYSTP

>O14
DIQLTQSPSSLSASVGDRVTITCRVSQGISSYLNWYRQKPGKVPKLLIYSASNLQSGVPS
RFSGSGSGTDFTLTISSLQPEDVATYYGQRTYNAPP

>O18
DIQMTQSPSSLSASVGDRVTITCQASQDISNYLNWYQQKPGKAPKLLIYDASNLETGVPS
RFSGSGSGTDFTFTISSLQPEDIATYYCQQYDNLP

>O2
DIQMTQSPSSLSASVGDRVTITCRASQSISSYLNWYQQKPGKAPKLLIYAASSLQSGVPS
RFSGSGSGTDFTLTISSLQPEDFATYYCQQSYSTP

>O4
DIQLTQSPSSLSASVGDRVTITCRVSQGISSYLNWYRQKPGKVPKLLIYSASNLQSGVPS
RFSGSGSGTDFTLTISSLQPEDVATYYGQRTYNAPP

>O8
DIQMTQSPSSLSASVGDRVTITCQASQDISNYLNWYQQKPGKAPKLLIYDASNLETGVPS
RFSGSGSGTDFTFTISSLQPEDIATYYCQQYDNLP

Exhibit D

The following sequences are taken from a study that has sequenced the entire human lambda gene locus (Kawasaki K. et al, 1997).

Total number of sequences: 36

>V1-11
QSVLTQPPSVSEAPRQRVTISCSGSSSNIGNNAVNWYQQLPGKAPKLLIYYDDLLPSGVS
DRFSGSKSGTSASLAISGLQSEDEADYYCAAWDDSLNGP

>V1-13
QSVLTQPPSVSGAPGQRVTISCTGSSSNIGAGYDVHWYQQLPGTAPKLLIYGNSNRPSGV
PDRFSGSKSGTSASLAITGLQAEDEADYYCQSYDSSLSGS

>V1-16
QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNTVNWYQQLPGTAPKLLIYSNNQRPSGVP
DRFSGSKSGTSASLAISGLQSEDEADYYCAAWDDSLNGP

>V1-17
QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNYVYWYQQLPGTAPKLLIYSNNQRPSGVP
DRFSGSKSGTSASLAISGLRSEDEADYYCAAWDDSLSGP

>V1-18
QSVLTQPPSVSGAPGQRVTISCTGSSSNIGAGYVVHWYQQLPGTAPKLLIYGNSNRPSGV
PDQFSGSKSGTSASLAITGLQSEDEADYYCKAWDNSLNA

>V1-19
QSVLTQPPSVSAAPGQKVTISCSGSSSNIGNNYVSWYQQLPGTAPKLLIYDNNKRPSGIP
DRFSGSKSGTSATLGITGLQTGDEADYYCGTWDSSLSAG

>V1-2
QSALTQPPSASGSPGQSVTISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYEVSKRPSGV
PDRFSGSKSGNTASLTVSGLQAEDEADYYCSSYAGSNNF

>V1-20
QAGLTQPPSVSKGLRQTATLTCTGNSNIVGNQGAAWLQQHQGHPPKLLSYRNNNRPSGIS
ERFSASRSGNTASLTITGLQPEDEADYYCSALDSSLSA

>V1-22
NFMLTQPHSVSESPGKTVTISCTRSSGSIASNYVQWYQQRPGSSPTTVIYEDNQRPSGVP
DRFSGSIDSSSNSASLTISGLKTEDEADYYCQSYDSSN

>V1-3
QSALTQPRSVSGSPGQSVTISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYDVSKRPSGV
PDRFSGSKSGNTASLTISGLQAEDEADYYCCSYAGSYTF

>V1-4
QSALTQPASVSGSPGQSITISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYEVSNRPSGV
SNRFSGSKSGNTASLTISGLQAEDEADYYCSSYTSSSTL

>V1-5
QSALTQPPSVSGSPGQSVTISCTGTSSDVGSYNRVSWYQQPPGTAPKLMIYEVSNRPSGV
PDRFSGSKSGNTASLTISGLQAEDEADYYCSLYTSSSTF

>V1-7
QSALTQPASVSGSPGQSITISCTGTSSDVGSYNLVSWYQQHPGKAPKLMIYEGSKRPSGV
SNRFSGSKSGNTASLTISGLQAEDEADYYCCSYAGSSTF

>V1-9
QSALTQPPFVSGAPGQSVTISCTGTSSDVGDYDHVFWYQKRLSTTSRLLIYNVNTRPSGI
SDLFSGSKSGNMASLTISGLKSEVEANYHCSLYSSSYTF

>V2-1
SYELTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRPSGIPER
FSGSNSGNTATLTISGTQAMDEADYYCQAWDSSTA

>V2-11
SYELTQPPSVSVSLGQMARITCSGEALPKKYAYWYQQKPGQFPVLVIYKDSERPSGIPER
FSGSSSGTIVTLTISGVQAEDEADYYCLSADSSGTYP

>V2-13
SSELTQDPAVSVALGQTVRITCQGDSLRSYYASWYQQKPGQAPVLVIYGKNNRPSGIPDR
FSGSSSGNTASLTITGAQAEDEADYYCNSRDSSGNHL

>V2-14
SYVLTQPPSVSVAPGQTARITCGGNNIGSKSVHWYQQKPGQAPVLVVYDDSDRPSGIPER
FSGSNSGNTATLTISRVEAGDEADYYCQVWDSSSDHP

>V2-15
SYELTQLPSVSVSPGQTARITCSGDVLGENYADWYQQKPGQAPELVIYEDSERYPGIPER
FSGSTSGNTTTLTISRVLTEDEADYYCLSGDEDNP

>V2-17
SYELTQPPSVSVSPGQTARITCSGDALPKQYAYWYQQKPGQAPVLVIYKDSERPSGIPER
FSGSSSGTTVTLTISGVQAEDEADYYCQSADSSGTYP

>V2-19
SYELTQPSSVSVSPGQTARITCSGDVLAKKYARWFQQKPGQAPVLVIYKDSERPSGIPER
FSGSSSGTTVTLTISGAQVEDEADYYCYSAADNNL

>V2-6
SYELTQPLSVSVALGQTARITCGGNNIGSKNVHWYQQKPGQAPVLVIYRDSNRPSGIPER
FSGSNSGNTATLTISRAQAGDEADYYCQVWDSSTA

>V2-7
SYELTQPPSVSVSPGQTARITCSGDALPKKYAYWYQQKSGQAPVLVIYEDSKRPSGIPER
FSGSSSGTMATLTISGAQVEDEADYYCYSTDSSGNH

>V2-8
SYELTQPHSVSVATAQMARITCGGNNIGSKAVHWYQQKPGQDPVLVIYSDSNRPSGIPER
FSGSNPGNTATLTISRIEAGDEADYYCQVWDSSSDHP

>V3-2
QTVVTQEPSLTVSPGGTVTLTCASSTGAVTSGYYPNWFQQKPGQAPRALIYSTSNKHSWT
PARFSGSLLGGKAALTLSGVQPEDEAEYYCLLYYGGAQ

>V3-3
QAVVTQEPSLTVSPGGTVTLTCGSSTGAVTSGHYPYWFQQKPGQAPRTLIYDTSNKHSWT
PARFSGSLLGGKAALTLLGAQPEDEAEYYCLLSYSGAR

>V3-4
QTVVTQEPSFSVSPGGTVTLTCGLSSGSVSTSYYPSWYQQTPGQAPRTLIYSTNTRSSGV
PDRFSGSILGNKAALTITGAQADDESDYYCVLYMGSGIS

>V4-1
QPVLTQPPSSSASPGESARLTCTLPSDINVGSYNIYWYQQKPGSPPRYLLYYYSDSDKGQ
GSGVPSRFSGSKDASANTGILLISGLQSEDEADYYCMIWPSNAS

>V4-2
QAVLTQPSSLSASPGASASLTCTLRSGINVGTYRIYWYQQKPGSPPQYLLRYKSDSDKQQ
GSGVPSRFSGSKDASANAGILLISGLQSEDEADYYCMIWHSSAS

>V4-3
QPVLTQPTSLSASPGASARLTCTLRSGINLGSYRIFWYQQKPESPPRYLLSYYSDSSKHQ
GSGVPSRFSGSKDASSNAGILVISGLQSEDEADYYCMIWHSSAS

>V4-4
QPVLTQPSSHSASSGASVRLTCMLSSGFSVGDFWIRWYQQKPGNPPRYLLYYHSDSNKGQ
GSGVPSRFSGSNDASANAGILRISGLQPEDEADYYCGTWHSNSKT

>V4-6
RPVLTQPPSLSASPGATARLPCTLSSDLSVGGKNMFWYQQKPGSSPRLFLYHYSDSDKQL
GPGVPSRVSGSKETSSNTAFLLISGLQPEDEADYYCQVYESSAN

>V5-1
LPVLTQPPSASALLGASIKLTCTLSSEHSTYTIEWYQQRPGRSPQYIMKVKSDGSHSKGD
GIPDRFMGSSSGADRYLTFSNLQSDDEAEYHCGESHTIDGQVG

>V5-2
QPVLTQPPSASASLGASVTLTCTLSSGYSNYKVDWYQQRPGKGPRFVMRVGTGGIVGSKG
DGIPDRFSVLGSGLNRYLTIKNIQEEDESDYHCGADHGSGSNFV

>V5-4
QPVLTQSSSASASLGSSVKLTCTLSSGHSSYIIAWHQQQPGKAPRYLMKLEGSGSYNKGS
GVPDRFSGSSSGADRYLTISNLQFEDEADYYCETWDSNT

>V5-6
QLVLTQSPSASASLGASVKLTCTLSSGHSSYAIAWHQQQPEKGPRYLMKLNSDGSHSKGD
GIPDRFSGSSSGAERYLTISSLQSEDEADYYCQTWGTG

Exhibit F

214–218 Nucleic Acids Research, 2000, Vol. 28, No. 1

Kabat Database and its applications: 30 years after the first variability plot

George Johnson and Tai Te Wu*

Departments of Biochemistry, Molecular Biology and Cell Biology, and of Biomedical Engineering, Northwestern University, Evanston, IL 60208, USA

Received July 23, 1999; Revised and Accepted October 13, 1999

ABSTRACT

The Kabat Database was initially started in 1970 to determine the combining site of antibodies based on the available amino acid sequences at that time. Bence Jones proteins, mostly from human, were aligned, using the now-known Kabat numbering system, and a quantitative measure, variability, was calculated for every position. Three peaks, at positions 24–34, 50–56 and 89–97, were identified and proposed to form the complementarity determining regions (CDR) of light chains. Subsequently, antibody heavy chain amino acid sequences were also aligned using a different numbering system, since the locations of their CDRs (31–35B, 50–65 and 95–102) are different from those of the light chains. CDRL1 starts right after the first invariant Cys 23 of light chains, while CDRH1 is eight amino acid residues away from the first invariant Cys 22 of heavy chains. During the past 30 years, the Kabat database has grown to include nucleotide sequences, sequences of T cell receptors for antigens (TCR), major histocompatibility complex (MHC) class I and II molecules and other proteins of immunological interest. It has been used extensively by immunologists to derive useful structural and functional information from the primary sequences of these proteins. An overall view of the Kabat Database and its various applications are summarized here. The Kabat Database is freely available at http://immuno.bme.nwu.edu

INTRODUCTION

The purpose of maintaining the Kabat Database of aligned sequences of proteins of immunological interest, in our opinion, is to provide useful correlations between structure and function for this special group of proteins from their nucleotide and amino acid sequences to their tertiary structures (1). These sequences are thus aligned with the ultimate aim of understanding how these proteins are folded and how they can perform their biological functions. We include only coding region sequences that have been published. In some cases, only the amino acid sequences were published, while the corresponding nucleotide sequences were deposited in GenBank. All stored sequences were then printed out and checked visually against available published sequences. We routinely survey for possible new sequences in journals in our libraries, Medline entries, cross-references from other papers, and author notification; however, we may still miss some sequences. GenBank, on the other hand, contains a substantial number of unpublished sequences. If there are doubts about these sequences or their annotations, please refer to the original papers. The Kabat numbering systems (see the Introduction of 2) for antibody light and heavy chains, for TCR alpha and beta chains, etc., go hand-in-hand with variability calculations. The locations of the CDRs are the theoretically derived positions which can be verified experimentally. Indeed, from the first antigen–antibody Fab complex (3) to the complexes of TCR, processed peptide and MHC class I molecule (4,5), it has been realized that alignment of amino acid sequences and variability calculations can be of utmost importance in understanding how these important macromolecules function biologically. Due to the rapid development of genetic and protein engineering methods, mouse and rat antibodies have been humanized to treat human cancers, viral infections, etc (6). CDRs of selected rodent antibodies are cut out and glued onto human antibody frameworks to minimize rejection by human patients.

Our predicted CDRs are slightly different from Chothia's. A careful comparison can be found from a hyperlink on our website to 'Andrew's Antibody Page' (http://www.biochem.ucl.ac.uk/~martin/abs/index.html).

Massive amounts of sequence data are being continuously published in the scientific literature. It is imperative to collect and properly align the sequences so that they can be used by as many researchers in this field as possible. We have previously published five editions of these sequences (see the Introduction of 2). In 1991, the fifth edition (2) consisted of three volumes. Currently, the database is more than five times as large. As of September 29, 1999, the Kabat database contained 1 599 375 and 2 517 756 nt for antibody light and heavy chain variable regions, respectively, as compared to 272 244 and 418 962 nt in 1991. Total numbers of entries, amino acids and bases of other categories of sequences can be obtained by using the 'Current Counts' hyperlink on our website. The collection is available on our website (http://www.immuno.bme.nwu.edu) which is free due to the generous support by various research grants from NIH since 1970.

Finally, numerous scientific papers have cited our database, quoting our fourth edition (7), fifth edition (2), or one of our more recent papers (8). On our part, we have been analyzing

*To whom correspondence should be addressed. Tel: +1 847 491 7849; Fax: +1 847 491 4928; Email: t-wu@nwu.edu

the Kabat Database during the past few years with reference to the total numbers of antibody and TCR V-genes, possible evolutionary selection processes, importance of antibody CDRH3s as related to their fine specificities, etc.

KABAT DATABASE

The Kabat Database may be accessed for searching, sequence retrieval and analysis by a few different methods: electronic mail, WWW and ftp. The electronic mail interface has been available since 1993, the WWW interface since 1995 and various formats of the database in electronic format for nearly a decade (8). Our data formats, searching tools, output formats and database structures have gradually been adopted by other immunological databases and interfaces.

Electronic mail interface

An electronic mail interface (seqhunt2@immuno.bme.nwu.edu) provides a non-interactive method for searching and sequence retrieval (9). Sending mail to the server address with the single word 'help' (no quotes) in the message body returns instructions for using the server.

All sequences classes are searchable and returnable. The query format allows making AND/OR/NOT constructed restrictions on the database and amino acid and nucleotide sequence pattern matching with allowable differences. Requests are processed as they are received and depending on the network traffic, take ~1–2 min to be searched and returned to the sender. The returned format is a fixed-line length record of 80 or fewer characters per line for ease in visual inspection and processing by user-written scripts or programs. The characters are plain text.

The query format for the sent request consists of two parts. The first part contains directives for the server to follow while the second part contains specifications of the search. Specification of the extent of data returned, the number of documents to return, starting document and whether plain ASCII text or PostScript should be used in the return format may be entered. Further, one can direct the server to return a distribution, the variability or unaligned raw data for the search specified.

The second part of the query contains the search restrictions on the database. Words separated by AND and OR may be used, as well as searching functions, like nucleotide/amino acid pattern matching and positional restriction matching.

There are basically three steps in translating and performing a search on the Kabat Database: generate the question or query, translate it into a format the server can recognize and decide on the output options desired of the returned matches. For example, if matches of mouse kappa light chains of anti-phosphorylcholine antibodies are desired, the query and restriction on the database would be:

Begin
@mouse and kappa and phosphorylcholine

The '@' before mouse tells the server that matches of the species mouse are desired, rather than searching through the entire database record for instances of the word 'mouse'. More complicated restrictions can be generated using parentheses for grouping and the minus sign '–' for NOT. Finding all rat and rabbit sequences which are not kappa light chains, and returning them as amino acid sequences in PostScript format would be constructed as:

PSAA
Begin
(rat and rabbit) and –kappa

Pattern matching is interpreted as the second part of an AND statement, such that finding all rat and rabbit sequences which are not kappa and contain the nucleotide pattern cagtacgtcag with three allowable mismatches, would be sent as:

Begin
(rat and rabbit) and –kappa [implicit AND]
#NM 3
cagtacgtcag

More examples of searching and output options may be found in the 'help' file returned from the server.

WWW interface

The WWW interface (8) to the Kabat Database: http://immuno.bme.nwu.edu contains searching and analysis tools as well as links to database download sites and other interesting databases. Most of the features found in the electronic mail interface are found in the WWW interface, as well as other tools. The WWW interface is more interactive than the Email and returns results faster, depending on the network traffic.

Searching and analysis tools

SeqhuntII. This grouping of programs allows searches through the annotations and sequence pattern matching of the amino acid and nucleotide sequence data with allowable mismatches. Like the Email server, restrictions on the database may be formulated as AND/OR/NOT constructs. Output extent, output format, maximum documents and starting document may be specified. Browsing of the output results in HTML format allows the user to view the database entries in an easy-to-read format. ASCII text may be selected as output for use in user-generated scripts and programs. PostScript generation allows for printing on a PostScript supporting printer. Sequence matching is returned aligned with the target sequence with nucleotide or amino acid differences from the database sequence displayed in a case change. Since the database contains only coding regions of genes and proteins, the query sequence should be a portion of the coding region being sought.

Variability. Variability and amino acid distributions of sequence groups may be generated for restrictions on the database. The variability plots are in PostScript format and may either be viewed on the screen with an appropriate PostScript viewer (e.g. GNU ghostscript or ghostview) or printed to a postscript-supporting printer. Plots for human and mouse TCR gamma and delta chain variable regions are shown in Figure 1. Scaling of the variability plots may be done allowing comparison of variability plots for different groupings of sequences. Distributions of the amino acids per position may be returned also, including the calculated variability for each position.

Sequence alignment. Alignment of user-entered coding regions of immunoglobulin light chains according to the Kabat numbering system can be performed. Because of the relatively few alignment options available for light chains, most sequences can be aligned. One can start with around 10 amino acid residues or 30 nt. There is no lower limit on the length of sequence to be matched. In some cases though, visual inspection and alignment is necessary, as is for heavy chain alignment,

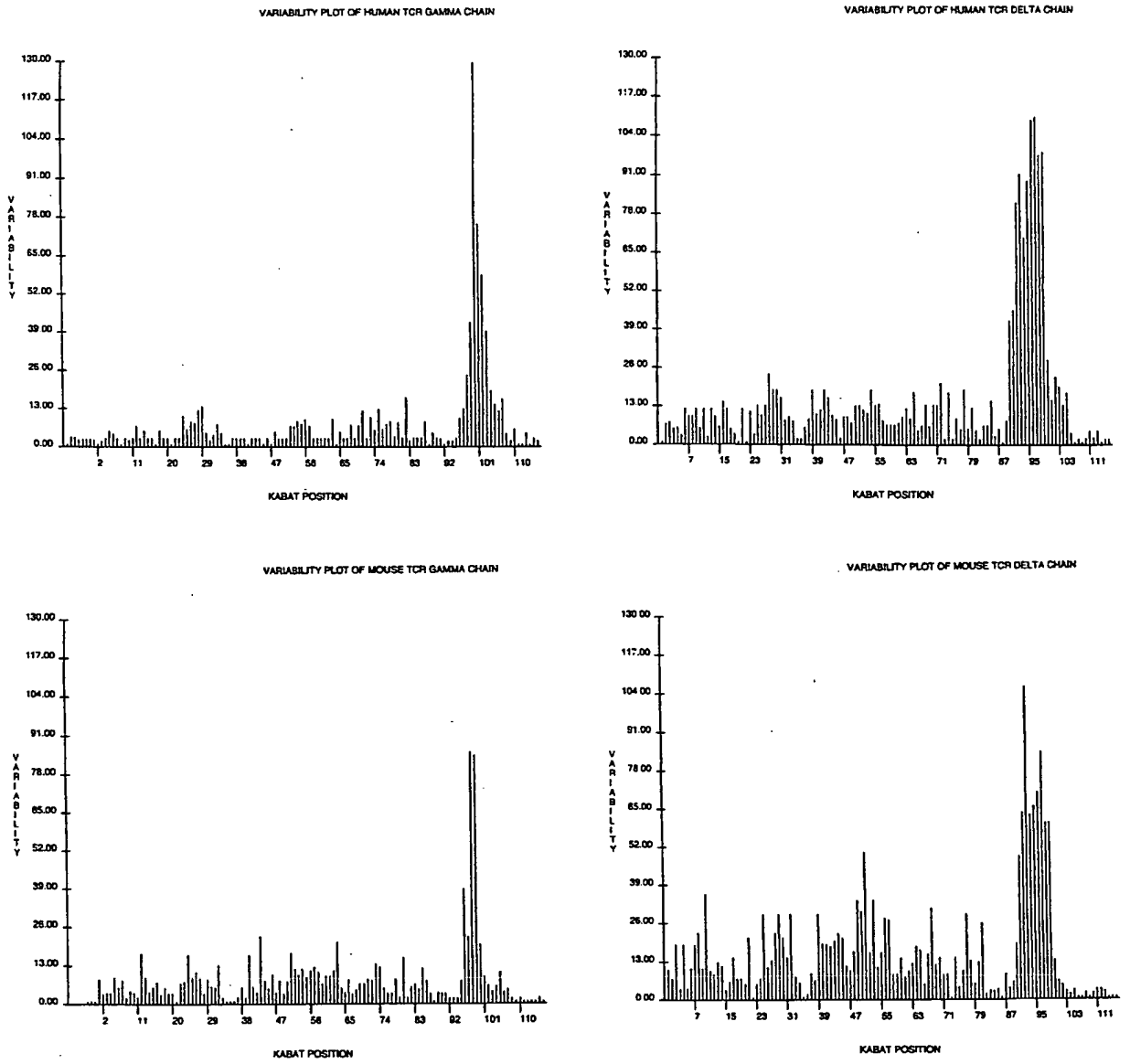


Figure 1. Variability plots for human and mouse TCR gamma and delta chain variable regions, using 377 human gamma, 1260 human delta, 297 mouse gamma and 461 mouse delta partial and complete sequences.

especially within the CDRH3 region, if additional codons or residues are inserted and denoted by '#'. If a suitable alignment counterpart from the database is not found for the target sequence, the user can contact us.

FTP. Various formats of the database are available for download from NCBI's repository under the directory 'kabat'. Currently active formats include a FASTA-like raw sequence format and the database's fixed length format of 80 or fewer characters per line and vertical alignment. Four main variations on the fixed length format exist to properly visually display single translations, pseudogene translations, J-minigenes and D-minigenes. Other immunological databases have adopted similar formats as exemplified by the three letter code amino acid translation followed by single letter code. A 'dump' version of the database is periodically updated which contains unlimited line length records more suitable for mass processing on unix-based systems.

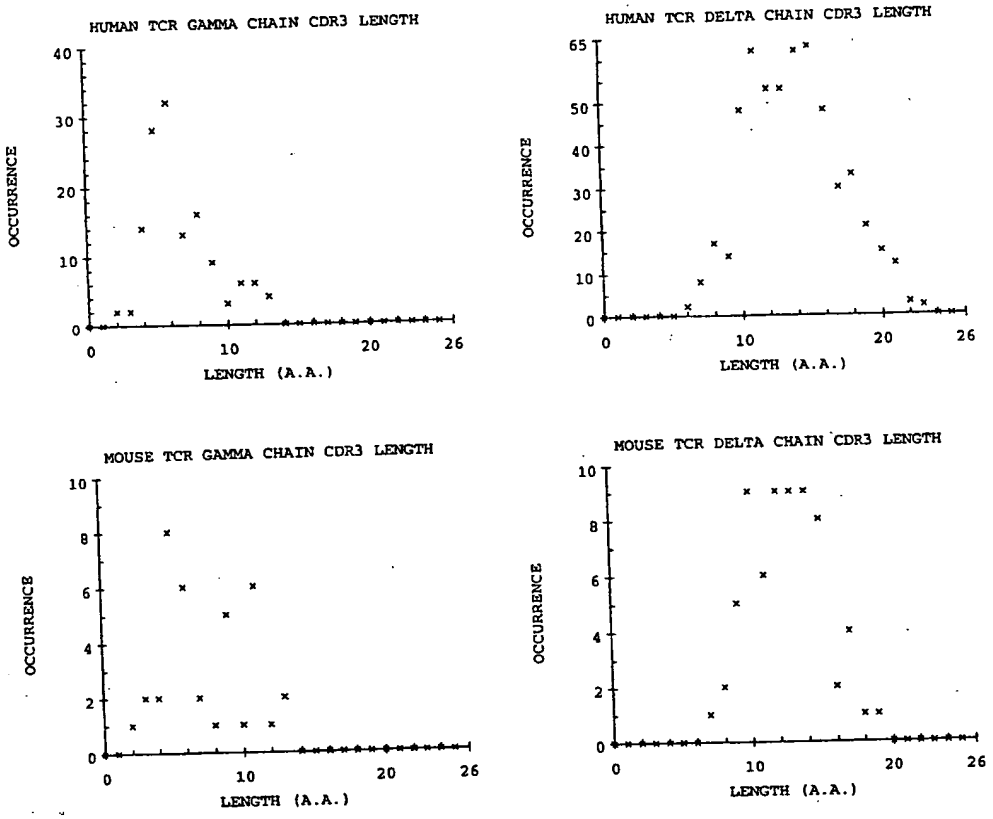


Figure 2. Length distributions of CDR3s of human and mouse TCR gamma and delta chains, based on 135 human gamma, 546 human delta, 37 mouse gamma and 66 mouse delta complete CDR3 sequences.

OTHER APPLICATIONS

As mentioned before, the Kabat Database was initially constructed for the purpose of identifying the antibody combining site (1). Starting from aligned amino acid sequences and using variability calculations, we have identified CDRs of antibody light and heavy chains, as well as those of TCRs. Such calculations can also provide useful predictions for MHC class I and II sequences (8), and to other aligned proteins sequences, e.g. HIV gp120, gp41, etc.

The importance of CDRH3 to confer fine specificity to antibodies was realized a few years ago (10). Furthermore, the unique CDRH3 nucleotide sequences have recently been used as a sensitive diagnostic test to detect residue B cell malignancies in cancer patients. Thus, many of these sequences have been determined. But most of them have been excluded from GenBank due to their relative short lengths. We have been meticulously collecting them, and realized the importance of their length distributions in antibodies of various specificities (11), and possible differences between CDRH3s of human and mouse (12). In the case of rabbit, the CDRH3s have less length variation than human and mouse. This may be compensated by the length variations of the CDRL3s (13).

The length variations of TCR alpha and beta chain CDR3s are very restricted (14). On the other hand, TCR gamma and delta chain CDR3s have more length variation, close to those of antibody heavy chains (Fig. 2). Whether they bind antigens directly is unclear.

During recent years, various research groups have decided to sequence the entire coding region of different antibody and TCR V-genes, in order to have an idea of their total numbers. On the other hand, we have discovered that pair-wise comparisons of V-gene nucleotide sequences in the Kabat Database provide very accurate estimations of their total numbers (15,16). In addition, such comparisons seem to suggest that antibody and TCR V-genes have evolved under different selective pressures (17). In the case of MHC class I sequences, comparison of their aligned sequences has elucidated a new mechanism of generating novel MHC class I molecules by random assortment of their a1 and a2 gene segments (18).

DISCUSSION

The Kabat Database has been around for 30 years. It has provided the immunology community a useful service, since it

not only is a sequence database but also incorporates vital aspects of the biology of the immune system. Various analytical methods have been developed to study the structure and function relations of proteins of immunological interest.
Electronic addresses:
http://immuno.bme.nwu.edu
seqhunt2@immuno.bme.nwu.edu
Citing the Kabat Database:
Authors using this database may cite this paper together with the electronic addresses.

ACKNOWLEDGEMENT

Supported in part by NIH Grant 5 R24 AI25616-10.

REFERENCES

1. Wu,T.T. and Kabat,E.A. (1970) *J. Exp. Med.*, 132, 211–250.
2. Kabat,E.A., Wu,T.T., Perry,H., Gottesman,K. and Foeller,C. (1991) *Sequences of Proteins of Immunological Interest*, Fifth Edition. NIH Publication No. 91-3242.
3. Amit,A.G., Mariuzza,R.A., Phillips,S.E.V. and Poljak,R.J. (1986) *Science*, 233, 747–753.
4. Garcia,K.C., Degano,M., Stanfield,R.L., Brunmark,A., Jackson,M.R., Peterson,P.A., Teyton,L. and Wilson,I.A. (1996) *Science*, 274, 209–219.
5. Garboczi,D.H., Ghosh,P., Utz,U., Fan,Q.R., Biddison,W.E. and Wiley,D.C. (1996) *Nature*, 384, 134–141.
6. Jones,P.T., Dear,P.H., Foote,J., Neuberger,M.S. and Winter,G. (1986) *Nature*, 321, 522–525.
7. Kabat,E.A., Wu,T.T., Reid-Miller,M., Perry,H. and Gottesman,K. (1987) *Sequences of Proteins of Immunological Interest*, Fourth Edition. US Govt. Printing Off. No. 165-492.
8. Johnson,G., Kabat,E.A. and Wu,T.T. (1996) In Herzenberg,L.A., Weir,W.M., Herzenberg,L.A. and Blackwell,C. (eds), *Weir's Handbook of Experimental Immunology I. Immunochemistry and Molecular Immunology*, Fifth Edition. Blackwell Science Inc., Cambridge, MA, pp. 6.1–6.21.
9. Johnson,G., Wu,T.T. and Kabat,E.A. (1995) In Paul,S. (ed.), *Antibody Engineering Protocols*. Humana Press, pp.1–15.
10. Kabat,E.A. and Wu,T.T. (1991) *J. Immunol.*, 147, 1709–1719.
11. Johnson,G. and Wu,T.T. (1998) *Int. Immunol.*, 10, 1801–1805.
12. Wu,T.T., Johnson,G. and Kabat,E.A. (1993) *Proteins*, 16, 1–7.
13. Sehgal,D., Johnson,G., Wu,T.T. and Mage,R.G. (1999) *Immunogenetics*, 50, 31–42.
14. Johnson,G. and Wu,T.T. (1999) *Immunol. Cell Biol.*, 77, 391–394.
15. Johnson,G. and Wu,T.T. (1997) *Genetics*, 145, 777–786.
16. Johnson,G. and Wu,T.T. (1997) *Immunol. Cell Biol.*,75, 580–583.
17. Johnson,G. and Wu,T.T. (1997) *J. Mol. Evol.*, 44, 253–257.
18. Johnson,G. and Wu,T.T. (1998) *Genetics*, 149, 1063–1067.

Exhibit 6

SEQUENCES OF PROTEINS OF IMMUNOLOGICAL INTEREST

FIFTH EDITION

Tabulation and Analysis of
Amino Acid and Nucleic Acid Sequences of Precursors,
V-Regions, C-Regions, J-Chain, T-Cell Receptors for Antigen,
T-Cell Surface Antigens, β_2-Microglobulins,
Major Histocompatibility Antigens, Thy-1, Complement,
C-Reactive Protein, Thymopoietin, Integrins, Post-gamma Globulin,
α_2-Macroglobulins, and Other Related Proteins

1991

Elvin A. Kabat[*], *Tai Te Wu*[†], *Harold M. Perry*[†],
Kay S. Gottesman[‡], *and Carl Foeller*[†]

[*]Depts. of Microbiology, Genetics and Development, and Neurology, Cancer Center/Institute of Cancer Research, College of Physicians and Surgeons, Columbia University, New York, NY 10032 and the National Institute of Allergy and Infectious Diseases, and the Office of the Director National Institutes of Health, Bethesda, MD 20892.

[†]Depts. of Biochemistry, Molecular Biology, and Cell Biology, and Engineering Sciences and Applied Mathematics and Biomedical Engineering, Northwestern University, Evanston, IL 60208 and the Cancer Center, Northwestern University Medical School, Chicago, IL 60611

[†]BBN Systems and Technologies, 10 Moulton Street, Cambridge, MA 02138

[‡]Formerly with BBN. Present address Laboratory for Applied Research in Academic Information, William H. Welch Medical Library, The Johns Hopkins University, Baltimore, Md 21205

The collection and maintenance of this data base is sponsored under grant 5R01 AI-125616 to E.A. Kabat of Columbia University by the following components of the National Institutes of

U.S. DEPARTMENT OF HEALTH
AND HUMAN SERVICES

Public Health Service
National Institutes of Health

NIH Publication No. 91-3242

TABLE I

Amino Acid Residues Associated with Framework (FR) and Complementarity Determining Regions (CDR) of the Variable Domains of Immunoglobulin Light (V_L) and Heavy (V_H) Chains

Segment	Light Chain	Heavy Chain
FR1	1-23 (with an occasional residue at 0, and a deletion at 10 in V_λ chains)	1-30 (with an occasional residue at 0)
CDR1	24-34 (with possible insertions numbered as 27A,B,C,D,E,F)	31-35 (with possible insertions numbered as 35A,B)
FR2[a]	35-49[a]	36-49
CDR2	50-56	50-65 (with possible insertions numbered as 52A,B,C)[b]
FR3	57-88	66-94 (with possible insertions numbered as 82A,B,C)
CDR3	89-97 (with possible insertions numbered as 95A,B,C,D,E,F)	95-102 (with possible insertions numbered as 100A,B,C,D,E,F,G,H,I,J,K)
FR4	98-107 (with a possible insertion numbered as 106A)	103-113

[a] Five Basilea rabbits (λ) immunized with type II pneumococci and which produced anti-type II pneumococcal polysaccharide had Met at position 48 and an insertion of four amino acid residues between positions 48 and 49; in four of the five the sequence was Glu, Leu, Lys, Ser and the fifth was Trp, Leu, Arg, Lys (53,54,63,64); the others were not sequenced at these positions (for references see table of rabbit λ amino acid sequences.)

[b] In the rabbit, Mage et al. (65) consider position 65 in V_H to be in FR3, since it is allotype related.

Exhibit H

RODGER G. SMITH, Ph.D.
3336 Eclipse Drive
Jefferson, MD 21755
301-834-7111 (Home)
240-314-4400 x 3497 (Office)
rodger_smith@hgsi.com

RESEARCH SCIENTIST

QUALIFICATIONS

- Ph.D. in Molecular Biology with 15 years of industrial laboratory experience.
- □ench scientist with expertise in phage antibody display technology, therapeutic antibody generation and characterization, in vitro assay development, protein purification and gene cloning and expression.
- Laboratory and project management experience.

PROFESSIONAL EXPERIENCE

Human Genome Sciences, Inc., Rockville, MD

Senior Scientist I, CoGenesys, Lead Identification **04/2005-present**

- Managed group involved in all aspects of the generation and characterization of fully human therapeutic antibodies to a variety of novel protein targets.

Senior Scientist I, Antibody Development **08/2002-03/2005**

- Lead scientist for project involved in the discovery and development of therapeutic antibodies to protective antigens from anthrax and plague causing organisms.
- Developed cloning protocols and selection techniques for affinity maturation of scFv's by phage display.
- Generated large panels of specific binding scFv's to a number of soluble and membrane-associated therapeutic target proteins.
- Developed a variety of electrochemiluminescent-based and cell-based assays to identify and characterize lead candidate therapeutic scFv's and/or IgG's.
- Project manager for outside collaboration focused on the discovery of therapeutic antibodies to 7TM chemokine receptors using human Fab phage display libraries.

Scientist, Antibody Development **04/2000-07/2002**

- Developed expression constructs and purification methods for producing homo-trimeric and heterotrimeric forms of BlyS and APRIL.
- Developed novel electrochemiluminescent-based assays to evaluate specificity and affinity of antibodies to a G protein-coupled receptor.
- Engineered mammalian antibody expression vectors to streamline cloning of V-domains and enhance expression of IgG.
- Developed methods for epitope mapping of recombinant antibodies using peptide phage display technology.

- Generated assay reagents and developed assay protocols for evaluating immunogenicity and pK of therapeutic antibodies.
- Generated a panel of fully human IgG's that blocked the catalytic activity of a novel angiotensin converting enzyme.

Bioveris Corp. (formerly IGEN International, Inc.), Gaithersburg, MD

Scientist 3, Assay Development **01/1999 – 04/2000**

- Manager for laboratory focused on the production, purification and validation of antibody and protein reagents used in diagnostic assay development.
- Developed and produced diagnostic and therapeutic antibody reagents on a contract basis using a large human repertoire phage antibody display library.
- Engineered a phage antibody display vector for production of novel binding reagents such as bivalent single-chain antibodies.

Scientist 1 and 2, Molecular Biology / Molecular Engineering **1990 to 1998**

- Principal Investigator (Phase I SBIR) for project focused on constructing a large human repertoire phage antibody display library. Responsibilities included project management, bench research and report writing.
- Optimized primer design, PCR conditions and cDNA cloning methods for generating large human and murine antibody repertoire phage display libraries.
- Coordinated research collaboration involving application of phage antibody display to isolate human and mouse single-chain antibodies with catalytic activity.
- Developed humanized single-chain and bivalent single-chain antibodies for tumor targeting and prodrug activation.
- Developed a bacterial cloning and expression system for high-level production of single-chain antibodies.
- Performed site-directed mutagenesis of an antibody enzyme (abzyme) to delineate catalytic residues.

SUMMARY OF LABORATORY EXPERTISE

Recombinant DNA methods:
mRNA isolation, RT-PCR, 5' RACE, cDNA cloning, vector construction, phage display library construction, site-directed mutagenesis, Northern and Southern hybridization, DNA sequencing, gene transfection, DNA replication assays, DNA diagnostic assays
Protein methods:
Protein Purification (FPLC and BioCad instrumentation), BIAcore analysis, antibody affinity measurements, immunoassay development, SDS-PAGE and Western blotting
Other skills:
Phage display technology, hybridoma production and screening, cell transfection and tissue culture, large scale bacterial fermentation and protein production

EDUCATION

1989 - Ph.D. in Microbiology (Molecular Biology), University of Illinois at Urbana-Champaign. Thesis research in the laboratory of Dr. Edward Voss studying the molecular basis of autoimmunity and anti-DNA autoantibodies in a murine model.

1984 - M.S. in Microbiology, University of Illinois at Urbana-Champaign.
Thesis research in the laboratory of Dr. John Scott studying yeast DNA replication enzymes associated with the chromatin of a freely replicating yeast DNA plasmid.

1981- B.S. in Microbiology, Pennsylvania State University, University Park. Cooperative Study for one year in the Microbiology Research Department at Hershey Foods, Inc., Hershey, PA.

GRANTS AND AWARDS

Phase I Small Business Innovation Research (SBIR) grant ($100,000).
Project Title-Construction of a Large Semi-Synthetic Human Phage Antibody Display Library, 1997. Granting Agency: Dept. of the Army.

U.S. Public Health Service Traineeship, 1984-1987, University of Illinois, awarded for academic achievement.

Clark Microbiology Award, 1983, 1988 and 1989, University of Illinois, for excellence in teaching.

ISSUED PATENTS

Reaction-Based Selection for Expression of and Concentration of Catalytic Moieties
US Patent Number 6,121,007 Date: Sept. 19, 2000 and 6,177,270B1 Date: Jan. 23, 2001

Cycling DNA/RNA Amplification Electrochemiluminescent Assay
US Patent Number 6,048,687 Date: April 11, 2000

The Isolation and Production of Catalytic Antibodies Using Phage Technology
European Patent Number P09002EPO Date: March 14, 2000

Prodrugs Activated by Targeted Catalytic Proteins
US Patent Number 6,258,360 B1 Date: July 10, 2001

Several Additional Patents Pending

PUBLICATIONS AND RECENT ABSTRACTS

1. Baker, K., Edwards, B., Main, H., Choi, G., Wager, R., Halpern, W., Lappin, P., Riccobene, T., Abramian, D., Sekut, L., Sturm, B., Poortman, C., Minter R., Dobson, C., Williams, E., Carmen, S., **Smith, R.**, Roschke, V., Hilbert, D., Vaughan, T., Albert, V. 2003. Generation and characterization of LymphoStat-B, a human monoclonal antibody that antagonizes the bioactivities of B lymphocyte stimulator. *Arthritis Rheum.* **48** (11), 3253-3265.

2. Huang, L., Sexton, D., Skogerson, K., Devlin, M., **Smith, R.,** Sanyal, I., Parry, T., Kent, R., Enright, J., Wu, Q., Conley, G., DeOliveira, D., Morganelli, L., Ducar, M., Wescott, C. and Ladner, R. 2003. Novel Peptide Inhibitors of Angiotensin-converting Enzyme 2. J. Biol. Chem. 278: 15532 - 15540.

3. Roschke, V., Sosnovtseva, S., Ward, C., Hong, J., **Smith, R.**, Albert, V., Stohl, W., Baker, K., Ullrich, S., Nardelli, B., Hilbert, D. and Migone, T. 2002. BlyS and APRIL form biologically active heterotrimers in patients with systemic immune-based rheumatic diseases. *Journal of Immunology* **169**, 4314-4321.

4. Abraham, R., Buxbaum, S., Link, J., **Smith, R.**, Venti, C. and Darsley, M. 1996. Determination of binding constants of diabodies directed against prostate-specific antigen using electrochemiluminescence-based immunoassays. *J. Mol. Recog* . **9**, 456 - 461.

5. Smith, R., Martin, M., Sanchez, R. and Kenten, J. 1995. Cloning and bacterial expression of an esterolytic sFv. *Meth. in Mol. Bio.* **51**, Antibody Engineering Protocols, S. Paul, editor, 297 - 317.

6. Abraham, R., Buxbaum, S., Link, J., **Smith, R.**, Venti, C. and Darsley, M. 1995. Screening and kinetic analysis of recombinant anti-CEA antibody fragments. *J. Imm. Meth.* **183**, 119 -125.

7. McCafferty, J., Fitzgerald, K. J., Earnshaw, J., Chiswell, D. J., Link, J., **Smith, R.** and Kenten, J. 1994. Selection and rapid purification of murine antibody fragments that bind a transition-state analog by phage display. *ABAB* 47, 157-174.

8. Gulliver, G., Bedzyk, W., **Smith, R.**, Bode, S., Tetin, S. and Voss, E. 1994. Conversion of an anti-single-stranded DNA active site to an anti-fluorescein active site through heavy chain complementarity determining region transplantation. *JBC* **269**, 7934.

9. Angeles, T., **Smith, R.,** Darsley, M., Sugasawara, R., Sanchez, R., Kenten, J., Schultz, P. and Martin, M. 1993. Isoabzymes: Structurally and mechanistically similar catalytic antibodies from the same immunization. *Biochemistry* **32**, 12128 -12135.

10. Kenten, J. and **Smith, R.** 1992. Catalytic antibodies from production to application. *Current Opinion in Therapeutic Patents* **2,** 669 -677.

11. Smith, R. and Voss, E. 1989. Variable region primary structures of monoclonal anti-DNA autoantibodies from NZB/NZW F1 mice. *Molecular Immunology* **27,** 463 -470.

12. Smith, R., Ballard, D., Blier, P., Pace, P., Bothwell, A., Herron, J., Edmundson, A. and Voss, E. 1989. Structural features of a murine monoclonal anti-ssDNA autoantibody. *J. Ind Inst. of Sci.* **69,** 25-46.

Poster Presentation at ICAAC, Sept. 13, 2003

Selection of Potent Neutralizing Human Monoclonal Antibodies to Protective Antigen of *Bacillus anthracis*. X. Zhang, J. Askins, R. Fleming, B. Sturm, C. Poortman, P. Viriassov, B. Peterson, M. Flynn, Y. Miao, D. Zukauskas, **R. Smith**, M. Laird, G. Choi. Human Genome Sciences, Inc. Rockville, MD

Scheduled for Poster Presentation at ASM, May 26, 2004

Development and Characterization of Fully Human anti-F1 Antibodies that Protect Against Lethal Challenge with *Yersinia pestis* in a Surrogate Mouse Model of Bubonic Plague. R.Fleming[1], D. Zukauskas[1], H. Heine[2], G. Andrews[2], S. Welkos[2], J. Adamovicz[2], M. Laird[1], G. Choi[1], **R. Smith**[1]

[1]Human Genome Sciences, Inc. Rockville, MD, [2]USAMRIID,Frederick, MD

Exhibit I

Fundamental Immunology, Third Edition,
edited by William E. Paul
Raven Press, Ltd., New York © 1993.

CHAPTER 9

Immunoglobulins

Structure and Function

Leon Carayannopoulos and J. Donald Capra

The antibody molecule plays the central role in humoral immunity by attaching to pathogens and then recruiting effector systems to destroy the invaders. In doing so, it embodies two antagonistic tendencies—diversity and commonality—since it must possess both a variable surface to recognize different foreign epitopes, and a constant surface that its own effector systems can recognize. Efforts to explain this duality continue to yield unexpected insights into immunology, biophysics, and molecular and evolutionary genetics—placing immunoglobulins among the most fruitful research subjects ever studied. In immunologic thought certainly, antibody research has been critical. For example, the paradoxical appearance of a highly variable N terminal coupled to a constant C regions in immunoglobulin chains led to the first articulation of the concept of "two genes, one polypeptide chain" (1) and eventually to class switching (see Chapters 10 and 22). Questions on the origin of binding-site diversity led to the discovery of VDJ recombination, a mechanism also operative in the T cell receptor locus (see Chapter 10). General biology also benefits from the immunoglobulin. For example, for all their complexity, antibodies are composed of segments derived from a single structural building block—the immunoglobulin domain. This domain is found in an ever-growing superfamily of proteins involved in immunologic, developmental, and neurologic cell recognition (2). In a different direction, several workers have generated antibody-binding sites that catalyze chemical reactions (3)—providing a new means of examining structure–function relationships in catalysis.

L. Carayannopoulos: Program in Biophysics and Medical Scientist Training Program, University of Texas, Southwestern Medical School, Dallas, TX 75235-9048.
J. D. Capra: Departments of Microbiology and Internal Medicine, University of Texas, Southwestern Medical School, Dallas, TX 75235-9048.

TABLE 2. *Numerical bounds of statistical and structural features of variable domains*

Statistical feature		Position	Corresponding structural loop
FR1	Heavy	1–30	
	Light	1–23	
CDR1	Heavy	31–35 (35a,b)	H1: 26–32
	Light	24–34 (27a–f)	L1: 26–33
FR2	Heavy	36–49	
	Light	35–49	
CDR2	Heavy	50–65 (52a–c)	H2: 53–55
	Light	50–56	L2: 50–52
FR3	Heavy	66–94 (82a–c)	
	Light	57–88	
CDR3	Heavy	95–102 (100a–k)	H3: 96–101
	Light	89–97 (95a–f)	L3: 91–96
FR4	Heavy	103–113	
	Light	98–107 (106a)	

CDR (complementarity-determining regions), FR (frameworks), H1–3 (heavy chain variable loops), and L1–3 (light chain variable loops) are numbered according to Kabat et al. (23). The numbering system serves to optimally align conserved residues. Length variations due to insertions/deletions during evolution or junctional diversification are labeled with letters. For example, heavy chain position 52 may represent up to three residues depending on the germline V gene employed. Data for the table were taken from Kabat et al. (23) and Chothia and Lesk (27).

cepts involved. The stereoviews depict the main-chain traces of two Fabs with the CDR α-carbons marked by dots and the heavy chain denoted by heavier lines. The V_L–V_H interface appears edge-on with strands G(3-3) of the light chain and C′ of the heavy chain closest to the viewer (reference to Fig. 4 may help orientation). The dots representing CDRs clearly cluster together at the top of the Fab to generate the binding site (Fig. 8B even shows the peptide ligand bound in the site). The crucial point to note is the very different CDR structure in the context of similar FR and C_H–C_L module backbones. The three-dimensional morphologies of V regions will be covered in more detail following further examination of the sequences that lead to binding site diversity.

V-Region Sequences

Variable region sequences do not randomly differ relative to each other—even within the CDRs. Analyses of hundreds of V regions reveal that the sequences naturally fall into a homology-based hierarchy directly related to the germline antibody gene loci. Members within a hierarchical group are more similar to each other than to all sequences from other groups; furthermore, similar sequences display a shared pattern of amino acid substitutions that serve as "membership badges" for the various classifications. Through examination of these linked-substitutions within and between species, an evolutionary history of V regions can be obtained. Of course, the oldest and most basic group is that of the V regions themselves, followed by the division into V_H, V_κ, and V_λ representing separate V-gene loci on different chromosomes. V_H, V_κ, and V_λ, in turn, split into V-region subgroups, or "families" (reviewed in ref. 25). The current requirements for membership in a family are based on nucleotide cross hybridization corresponding to 80% homology at the DNA level. At the protein level this translates to about 75% identity within families and less than 70% (usually 30–60%) between families. By this criterion, V_H splits into six families as do V_κ and V_λ. Table 3 shows the percent identity between representative members of the six human V_H families. Note how the comparisons of family members (boxed numbers) yield identity values over 80%. The split into six families antedates the primate–rodent divergence since mice possess analogous subgroups. Figure 9 aligns six sequences —two each from V_H, V_κ, and V_λ—to demonstrate linked substitutions. All six sequences display a V-specific motif in FR4 (W/F-G-X-G) (19); this provides a β-bulge in strand 3-3 crucial for correct dimerization. In turn, other patterns distinguish heavy from light chains. Note the V_H-specific sequence (G-L-E-W-hydrophobic) in the middle of FR2 as compared to the V_L-specific motif (P-hydrophilic-hydrophobic-L-hydrophobic) in the analogous location; these residues induce β-bulges in strand C′ also for V dimerization (26). More extensive patterns differentiate the families from each other, as shown in Fig. 10 for some κ and H family members. The importance of V-region classification reaches beyond academic or evolutionary interest. The splitting of families can be continued down to subfamilies and finally individual genes. Thus family extent and membership directly reflect the germline antibody repertoire of an individual. Since the V-region loci are known to be polymorphic, presence or absence of a given gene may influence the predisposition to autoimmunity (25) (see Chapter 30) and/or the prenatal "education" of the immune system through idiotypic networks (see Chapter 24). More generally, the various gene groups provide the universe of starting structures available to the humoral immune system except for CDR3. Correlation of these sequences with their corresponding structures provides not only insights to the limits of binding site variability, but also a means for semiempirical prediction of V-region structure based on the sequence (27,28).

Full sequence variability results from participation of the germline genes in the somatic diversification system described in Chapter 10. In this system, about 100 functional V genes present at each of the V_H, V_κ, and V_λ loci encode unique (FR1-CDR1-FR2-CDR2-FR3-5′ CDR3) segments while four to six "joining (J)" minigenes at each locus encode a unique (3′ CDR3-FR4) section. The heavy chain employs an additional minigene, termed "diversity (D)," to encode the central portion of CDR3; about 30 unique D genes exist between the V_H and J_H

Human Light Chain Sequences from Kabat Database with an Amino Acid at Position 0

Sequence name	Sequence
HBL-2'CL	AQSVLTQPPS-VSAAPGQKVTISCSGSNSN----IGNNYVSWYQQLPGRAPKLLIYDNNERPSGIPDRFSGSKSGTSATLAITGLQTGDEADYYCGTWDSSLNF----VVFGGGTKLTVLG--
HBL-3'CL	AQSVLTQAPS-VSGPPGQSVTVSCSGNTSN----IGTKTVDWYQHFPGVAPRLLIYSTSQRPSGVPDRFSGSRSGTSASLAISGLQSEDEADYYCATWDDSLNF----YVFGSGTTVIVLG--
MP9'CL	DSYQLTQPPS-VSGSPGQSITISCTGTSSDV---GSYNLVSWYQQHPGEAPKLIIYEVSKRPSGVSNRFSGSKSGNTASLTISGLQAEDEAEYYCCSYAADST-----VIFGGGTKLTVL---
RPMI8226'CL	ELSVLTQPAS-VNGSPGQLIIISCTGPSSDI---GDYQYISWYQQHPGKAPKLIIYDVDKRPSGVSNRFSGSKSGNTASLTISGLQAEDEADYYCSSYRGSALFE---VVFGGGTKVTVLRQP
MC116'CL	ELSVLTQPAS-VSGSPGQSVTISCTGTSSEV---GGYDYVSWYQQHPGKAPKLMIYEVTNGPSGVSNRFSLSKSGNTASLTISGLQAEDEADYYCSSYTSSRPL----VVFGGGTKLTVLGQP

References

HBL-3'CL,	RIBOLDI,P. et al., BLOOD 83: 2952-2961 (1994)
HBL-2'CL	RIBOLDI,P. et al., BLOOD 83: 2952-2961 (1994)
MP9'CL	ANDRIS,J.S. et al., MOL IMMUNOL 32: 1105-1122 (1995)
RPMI8226'CL	WATKINS,B.A. et al., SCAND J IMMUNOL 42: 442-448 (1995)
MC116'CL	WATKINS,B.A. et al., SCAND J IMMUNOL 42: 442-448 (1995)

Exhibit 5

Exhibit K

Two Acquired Immunodeficiency Syndrome-Associated Burkitt's Lymphomas Produce Specific Anti-i IgM Cold Agglutinins Using Somatically Mutated V_H4-21 Segments

By Piersandro Riboldi, Gianluca Gaidano, Edward W. Schettino, Thomas G. Steger, Daniel M. Knowles, Riccardo Dalla-Favera, and Paolo Casali

We analyzed the reactivity and the structure of the V_H and V_L segments of two IgM monoclonal antibodies (MoAbs) produced by spontaneously in vitro outgrowing cell lines, HBL-2 and HBL-3, established from two acquired immunodeficiency syndrome (AIDS) patients with Epstein-Barr virus (EBV)-negative Burkitt's lymphoma (BL). These B-cell clones were representative of the respective neoplastic parental clones, as determined by immunophenotypic and molecular genetic analysis. The IgM MoAbs were highly specific for the i determinant on red blood cells (cold agglutinins), but bound none of the other eight self and nine foreign antigens (Ags) tested, including those most commonly recognized by natural antibodies or autoantibodies. Structural analysis showed that the IgM MoAb V_H segment sequences were 93.5% and 84.2% identical with that of the germline V_H4-21 gene, which encodes the vast majority of cold agglutinins that are specific for the i/I carbohydrate Ag and are produced under chronic lymphoproliferative conditions. The HBL-2 MoAb V_H4-21 gene segment was juxtaposed with 20P3 and J_H6 genes and paired with a Vλ1 segment, the sequence of which was 95.5% identical to that of the germline Humlv117 gene; the HBL-3 MoAb V_H4-21 gene segment was juxtaposed with DXP'1 and J_H5 genes and paired with a Vλ1 segment, the sequence of which was 86.7% identical to that of the germline Humlv1L1 gene. The high degree of conservation of the V_H4-21 gene in the human population, the nature of the nucleotide differences in the expressed V_H4-21 segments, and the presence of nucleotide substitutions in the HBL-2 and HBL-3 IgM MoAb J_H and/or Jλ segments suggested that the MoAb V segments underwent a process of somatic hypermutation. This was formally shown in the HBL-3 MoAb V_H segment, by differentially targeted polymerase chain reaction amplification of the HBL-3 MoAb-producing cell genomic DNA. In addition, cloning and sequencing of the genomic DNA from fibroblasts of the same patient whose neoplastic B cells gave rise to the HBL-3 cell line yielded a germline copy of the V_H4-21 gene. Thus, the expression of V_H4-21 gene products may be involved in a self Ag-driven process of clonal B-cell expansion and selection associated with BL in these AIDS patients.

AUTOIMMUNE PHENOMENA can occur in association with several human B-cell disorders, such as cold agglutinin disease,[1,2] lymphoma,[3-6] and the B-cell expansion and hypergammaglobulinemia occurring in human immunodeficiency virus (HIV)-infected patients.[7-9] Although the precise role of different self antigens (Ags) in the B-cell clonal selection associated with the above pathologic conditions remains to be defined, circumstantial evidence for a role of self Ags in clonal expansion and selection in autoimmune humans and mice has been provided.[10-16] The crucial role of Ags in inducing clonal expansion and selection in the normal B-cell repertoire is well documented.[17-21] Recently, it has been suggested that Ag stimulation also plays a role in the B-cell expansion and selection preceding and/or associated with development of lymphomas of various histologic types.[5,6,22-24]

The assessment of a potential role for Ags in the clonal B-cell expansion and selection associated with lymphoma or leukemia entails, first of all, the definition of the specificity and of V_H and V_L segment structure of the tumor-derived antibody. The analysis of antibody specificity and V_H and V_L segment structure depends on the availability of a homogeneous in vitro growing and Ig-producing tumor cell population representative of the in vivo neoplastic clone. This is a critical requirement in view of the findings of oligoclonal or polyclonal, not necessarily neoplastic, B-cell populations accompanying the predominant neoplastic clone, as found particularly in bioptic specimens of Burkitt's lymphoma (BL) emerging in patients with acquired immunodeficiency syndrome (AIDS).[25,26] Possibly due to the improved treatment and longer survival rate, these patients display a 60-fold increased incidence, relative to that expected for the general population, of lymphomas, mainly of the Burkitt's type.[27]

We analyzed the Ag reactivity and the structure of the V_H and V_L segments of the IgM monoclonal antibodies (MoAbs) produced by two spontaneously in vitro outgrowing cell lines established from 2 AIDS patients with Epstein-Barr virus (EBV)-negative BL.[28] The absolute identity between the in vitro growing monoclonal cell lines and their respective in vivo neoplastic clones was established by immunophenotypic and molecular genetic analysis. The IgM MoAbs from both cell lines strongly bound to the i Ag on red blood cells (RBCs; cold agglutinins), but to none of the other self and foreign Ags tested. The structural correlate for such Ag-binding specificity was provided by segments encoded by V_H4-21 and Vλ1 genes in somatically mutated configuration.

From the Department of Pathology and Kaplan Comprehensive Cancer Center, New York University School of Medicine, New York, NY; and the Department of Pathology, College of Physicians and Surgeons, Columbia University, New York, NY.

Submitted November 2, 1993; accepted January 19, 1994.

Supported by the US Public Health Service Grants No. AR-40908, CA-16087, and CFAR 5P30 AI-27742 (to P.C.), and CA-37895 (to R.D.-F.) and the Ministry of Health, I.S.S. (Italy) AIDS Project 8206-07 1993 (to P.R.). G.G. was supported by a fellowship for AIDS research from the Ministry of Health, I.S.S. (Italy). P.C. is a Kaplan Cancer Scholar. This is publication 22 from The Jeanette Greenspan Laboratory for Cancer Research.

Address reprint requests to Paolo Casali, MD, Department of Pathology, MSB-599, New York University School of Medicine, 550 First Ave, New York, NY 10016.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. section 1734 solely to indicate this fact.

0006-4971/94/8310-0024$3.00/0

Thus, a process of selection by the i self Ags may have played a role in the B-cell clonal expansion preceding and/or associated with the development of BL in these AIDS patients.

MATERIALS AND METHODS

Generation and characterization of the monoclonal AIDS BL cell lines. The HBL-2 and HBL-3 MoAb-producing cell lines were established using B lymphocytes spontaneously outgrowing from two tumors histologically and immunophenotypically classified as small noncleaved cell lymphoma (SNCCL).[28] Both SNCCL arose in patients with AIDS. Immunophenotypic analysis was performed by fluorescence flow cytometry of isolated cells using a FACScan (Becton Dickinson Corp, Mountain View, CA) and a panel of labeled murine MoAbs, including those to CD19, HLA-DR, CD10, CD21, and CD5.[28] The clonality of the cell lines and their absolute relatedness to the tumors were determined by Ig gene rearrangement analysis using a J_H probe on *Hin*dIII, *Eco*RI, and *Bam*HI DNA digests.[29] The c-*myc* translocations were detected by cytogenetic analysis.[28] The status of the c-*myc* locus was analyzed by hybridization of *Eco*RI- and *Hin*dIII-digested DNA to the human c-*myc* probe MC413RC, representative of the third exon of the c-*myc* gene.[30] The presence of the EBV genome was investigated using a probe specific for the EBV genomic termini (5.2-kb *Bam*HI-*Eco*RI fragment isolated from the fused *Bam*HI terminal fragment NJ-het).[28] The presence of HIV sequences was investigated using the λ7A/2 probe on *Hin*dIII and *Sac* I DNA digests, and that of HTLV-I sequences was investigated using an HTLV-env probe on *Bam*HI and *Pst* I DNA digests.[28]

Analysis of the AIDS BL cell line-derived MoAb. The IgM MoAbs produced by the HBL-2 and HBL-3 cell lines were analyzed for their binding to polyclonal human IgG Fc fragment (Organon Teknika-Cappel, Malvern, PA); calf thymus DNA (Sigma Chemical Co, St Louis, MO); insulin (Sigma); human recombinant tumor necrosis factor-α (TNF-α), TNF-β, and interleukin-1β (IL-1β; BASF Biotech Corp, Cambridge, MA); human thyroglobulin; HIV-1 and cytomegalovirus (CMV) and parvovirus B19 recombinant glycoproteins; lipopolysaccharide (LPS) and β-galactosidase from *Escherichia coli* (Sigma); phosphorylcholine chloride (Sigma); *Pneumococcus* polysaccharides, including types 1, 3, and 4; and tetanus toxoid (Massachusetts Public Health Biological Laboratories, Jamaica Plain, MA), using specific enzyme-linked immunosorbent assay (ELISA) involving plates coated with 1 to 5 μg/mL of these Ags.[31-34] The IgM MoAb specific *i/I* blood group cold agglutinin activity was tested by hemagglutination using papain-treated and untreated group 0^+ umbilical cord or adult human erythrocytes in a 100 μL reaction volume.[35] Cold agglutinating titers were expressed as the smallest amount (nanograms) of IgM MoAb agglutinating 10^7 papain-treated human RBCs. Finally, the presence of the cross-reacting 9G4 idiotype on the IgM HBL-2 and HBL-3 MoAbs was analyzed using the appropriate IgG2a rat antibody.[1,2,36]

Cloning and sequencing of the MoAb V_H and Vλ genes. Poly(A)$^+$ RNA was isolated from the HBL-2 and HBL-3 MoAb-producing B-cell lines using the Micro Fast-Track mRNA isolation kit (Invitrogen Corp, San Diego, CA) and reverse-transcribed using Moloney murine leukemia virus (MMLV) reverse transcriptase (SuperScript RNase H-Reverse Transcriptase; GIBCO BRL Life Technologies, Gaithersburg, MD). cDNA (100 ng) was submitted to polymerase chain reaction (PCR) in a volume of 50 μL containing 200 μmol/L of each dNTP, 2.5 U of *Taq* polymerase (Perkin Elmer Corp, Norwalk, CT), and 10 pmol of each primer. The following sense degenerated primers encompassing a portion of the leader region of different V_H families plus an *Eco*RI site were used: ALT 1 (V_HI family) [5′ GGGAATTCATGGACTGGACTGGACCTGGAGG(AG)TC(CT)TCT(GT)C 3′]; ALT2 (V_HIII family) [5′ GGGAATTCATGGAG(CT)TTGGGCTGA(CG)CTGG(CG)TTT(CT)T 3′]; ALT4 (V_HIV family) [5′ GGGAATTCATGAA(AG)CA(TC)CTGTGGTTCTT(CT)(AC)T-(CT)CT(CG)C 3′]; and HI-3, an oligonucleotide primer encompassing a portion of the leader sequence of V_HIII family, [5′ TTGGGCTGTGCTGGGTTTTCCT 3′]. The antisense primer consisted of the reverse complement [5′ CCGAATTCAGCCGAGGGGGAAAAGGGTTT 3′] of a 21 nucleotide 5′ Cμ sequence plus an *Eco*RI site. The oligonucleotide primers specific for the 5′ portions of the Vλ chains were as follows: VλI[5′ ATG(GA)CC(TG)GCT(CT)CCCTCTCCTCCT 3′], and VλII-VI (VλII, VλIII, VλIV, VλVI chains) [5′ ATG(AG)C(CT)TGGACCC(CT)(AT)CTC(CT)(TG)(TG)TT 3′]. The antisense Cλ chain primer consisted of the [5′ TTGGCTTGAAGCTCCTCAGAGGA 3′] oligonucleotide. For V_H and Vλ gene amplification, 30 cycles of PCR were performed under denaturing, annealing, and extending conditions of 94°C (1 minute), 52°C (1 minute), and 72°C (2 minutes), respectively. PCR products were sized and isolated on low melting agarose gel and ligated into pCR 1000 plasmid vector (Invitrogen Corp). The ligation mixture was used to transform INVαF′ competent cells according to the manufacturer's protocol. Recombinant clones were selected according to the length of the insert and sequenced by the dideoxy chain termination method using the *Taq* Track Sequencing Kit (Promega Corp, Madison, WI). Each V_H or Vλ sequence was derived from the analysis of at least three independent clones. Differences in nucleotide sequences among different recombinant clones were observed in few cases (<0.001/base) and such variants were excluded from the sequence analysis. DNA sequences were analyzed using the software package of the Genetics Computer Group of the University of Wisconsin, version 7.1, and a Model 6000-410 VAX computer (Digital Equipment Corp, Marlboro, MA). DNA sequence identity searches were performed using the GenBank database and the FASTA method.[37]

Analysis of the putative germline IgV_H segment that gave rise to the expressed HBL-3 V_H gene. Genomic DNA was extracted from the monoclonal HBL-3 B cells and autologous fibroblasts obtained from the same bioptic sample used for the generation of the tumoral HBL-3 cell line. B-cell or fibroblast genomic DNA (100 ng) was supplemented with the appropriate sense and antisense oligonucleotide primers (10 pmol each). PCR amplification was performed in a 50 μL reaction volume using *Taq* polymerase under denaturing, annealing, and extending conditions of 94°C (1 minute), 60°C (1 minute), and 72°C (1 minute), respectively, for 30 cycles. The oligonucleotides used were as follows: (1) the sense V_H4-21 FR1 primer, encompassing a portion (residues 10 to 27) of the FR1 sequence of germline V_H4-21 gene[38] and differing in two nucleotides from the corresponding sequence of the expressed HBL-3 V_H gene [5′ CTACAGCAGTGGGGCGCA 3′]; (2) the sense HBL-3 leader primer, encompassing a portion of the leader sequence of the HBL-3 V_H gene and differing in one nucleotide (C instead of G at position −31) from the corresponding area of the germline V58 gene,[39] the member of the V_HIV gene family displaying the highest degree of identity with the V_H4-21 gene; and (3) the antisense V_H4-21 FR3 primer, consisting of the reverse complement [5′ GTGTCCGCGGCGGTCACAGA 3′] of a FR3 sequence (residues 250 to 269) shared by the expressed HBL-3 V_H gene and the germline V_H4-21 gene. Part of the amplified DNA was fractionated through a 1.5% agarose gel, transferred to a nylon membrane (Hybond; Amersham Life Sciences, Arlington Heights, IL) and hybridized at 48°C with the HBL-3 complementarity-determining region 1 (CDR1) oligonucleotide probe labeled with [γ-^{32}P] ATP (DuPont NEN Research Products, Boston, MA) by T4 polynucleotide kinase. The HBL-3 CDR1 oligonucleotide encompassed a FR1-CDR1 sequence [5′ GTTCCGTGATTACCTCTGGACA 3′] (residues 84 to 105) of the expressed HBL-3 V_H gene, differing in seven bases from that of the

Table 1. Features of the AIDS BLs

Tumor							MoAb			V_H Segment	
							Chains				
Cell Line	Source	Histology	EBV DNA	HIV DNA	HTLV-I DNA	c-*myc*	H	L	Ag Specificity	V_H Gene	Nucleotide (amino acid) Identity* (%)
HBL-2	Pleural fluid	SNCC	No	No	No	t(8;14)	μ	λ	i 4§	V_H4-21‖	93.5 (88.6)
HBL-3	Liver mass	SNCC	No	No	No	t(8;22)	μ	λ	i 0.6§	V_H4-21	84.2 (80.4)

* Compared with the genomic germline sequences.

† The gene 20P3 has been reported by Schroeder et al[47]; DXP'1 gene has been reported by Ichibara et al.[44]

‡ The complete sequences of the genomic germline V_L genes have been reported as follows: HumIv117 (Vλ1 subgroup),[15] and HumIv1L1 (Vλ1 subgroup).[51]

§ Smallest amount (in nanograms) of IgM MoAb agglutinating 10^7 papain-treated cord RBCs in a 100 μL reaction volume.

‖ The sequence of the genomic germline V_H4-21 gene has been reported by Sanz et al.[38]

¶ Expected numbers of R mutations calculated as reported in the Results.

corresponding area of the germline V_H4-21 gene. After hybridization, the membrane was washed twice with 2× SSC/0.1% sodium dodecyl sulfate (SDS) at room temperature for 30 minutes and twice with 1× SSC/0.1% SDS at 54°C for 30 minutes. Autoradiography was performed using Kodak XAR-5 film (Eastman Kodak Co, Rochester, NY). Other amplified DNA was inserted into the pCR II plasmid vector (Invitrogen Corp) for cloning and sequencing.

RESULTS

Characterization of the IgM MoAb-producing HBL-2 and HBL-3 cell lines. The features of the HBL-2 and HBL-3 cell lines established from 2 AIDS patients with SNCCL and their identity with the respective primary tumoral tissues have been reported[28] and are summarized in Table 1. Both tumor samples and respective cell lines expressed surface Ig μ and λ chains and the B-cell–restricted marker CD19. This was consistent with the B-cell origin of the cell lines and with the histologic diagnosis of SNCCL.[40] In addition, the surface expression of CD10 but not CD21 was consistent with the cell line phenotype of sporadic BL.[40] The monoclonality of the AIDS BL cell lines and their absolute relatedness to the respective tumors was formally established by the analysis of the Ig H chain gene rearrangements and that of the c-*myc* oncogene translocations. The Southern blotting analyses of *Eco*RI- and *Hin*dIII-digested DNA were consistent with two distinct patterns of c-*myc* activation. In the HBL-2 cells, the breakpoint was located 5′ (3 to 5 kb) to the c-*myc* first exon; in the HBL-3 cells, the breakpoint was within a ~4-kb region 3′ of the c-*myc* exon 3. Both HBL-2 and HBL-3 cells were negative for EBV, HIV, and HTLV-1 sequences and so were their respective original tumor cells.

The Ag-binding activity of the IgM MoAbs produced by the HBL-2 and HBL-3 cell lines. The HBL-2 and HBL-3 IgM MoAbs specifically bound to the i determinant on human erythrocytes, as shown by their strong agglutination of cord (smallest agglutinating doses: 4 and 0.6 ng/10^7 RBCs, respectively), but not adult (no agglutination by HBL-2 MoAb; HBL-3 MoAb smallest agglutinating dose, 38.8 ng/10^7 RBCs) papain-treated human RBCs. The specificity of the HBL-2 and HBL-3 IgM MoAbs was further strengthened by the MoAbs' failure to bind to any of the eight self and nine foreign Ags tested, including IgG Fc fragment, human thyroglobulin, ssDNA, phosphorylcholine, insulin, human recombinant TNF-α, human recombinant TNF-β, human recombinant IL-1β, β-galactosidase and LPS from *E coli*, tetanus toxoid, HIV-1, recombinant glycoproteins of CMV and Parvovirus B19, and *Pneumococcus* polysaccharides (Table 1). Binding to each of these Ags by HBL-2 and HBL-3 MoAbs yielded an absorbance of less than 0.05 at 492 nm; negative and positive controls were always less than 0.05 and more than 1.00, respectively. The HBL-3 but not the HBL-2 MoAb expressed the cross-reacting idiotype defined by the anti-idiotypic 9G4 MoAb.[1,2,36]

The HBL-2 and HBL-3 MoAb V_H segments. Figure 1 depicts the nucleotide (A) and deduced amino acid (B) sequences of the HBL-2 and HBL-3 IgM-MoAb V_H genes and that of the closest reported germline V_H gene. The differences between sequences are summarized in Table 1. The HBL-2 and HBL-3 MoAb V_H gene sequences were 93.5% and 84.2% identical, respectively, to that of V_H4-21 gene, a member of the V_HIV gene family.[38] Accurate inspection of the V_H4-21–related sequences available in the GenBank showed that the HBL-2 V_H gene sequence displayed an identical nucleotide difference at position 82, A instead of T, resulting in the same replacement of a Ser with a Thr, as found in the sequence of the V_H4-21 allele HumigvH4c.[41] Twelve of the 19 nucleotide differences displayed by the HBL-2 V_H gene sequence resulted in putative amino acid replacements, yielding replacement to silent (R:S) mutation ratios of 5.0 in the CDR and 1.1 in the framework regions (FR). Nineteen of the 46 nucleotide differences displayed by the HBL-3 V_H sequence resulted in putative amino acid replacements, yielding R:S mutation ratios of 2.2 in the CDR and 0.3 in the FR.

The HBL-2 and HBL-3 MoAb D and J_H genes. The nucleotide and deduced amino acid sequences of the HBL-2 and HBL-3 IgM MoAb D and J_H genes, and those of their closest germline D[42-46] and J_H genes[42] are depicted in Fig 1. The HBL-2 MoAb D gene sequence displayed some identity with that of the expressed fetal 20P3 D gene[47]; the HBL-3 MoAb D gene displayed a stretch of similarity to that of the

and Related Cold Agglutinin MoAbs

V_H Segment						Vλ1 Segment						
Nucleotide Differences in								Nucleotide Differences in				
CDR		FR					Nucleotide (amino acid) identity* (%)	CDR		FR		
R	S	R	S	D Gene†	J_H Gene	Vλ1 Gene‡		R	S	R	S	Jλ Gene
5 (3.2)¶	1	7 (10.1)¶	6	20P3	J_H6	Humlv117	95.5 (92.8)	5 (2.7)¶	4	2 (6.9)¶	2	Jλ2/Jλ3
11 (7.8)¶	5	8 (26.2)¶	22	DXP'1 (inverted)	J_H5	Humlv1L1	86.7 (80.6)	9 (8.0)¶	8	10 (20.8)¶	12	Jλ1

reverse complement of the germline DXP'1 gene, suggesting a possible inverted D gene joining origin (Fig 1C).[48] Both expressed D genes were flanked by untemplated nucleotide additions. The entire length of the D segment ranged from 13 nucleotides in HBL-2 to 36 nucleotides in HBL-3. The HBL-2 and HBL-3 MoAbs used truncated and mutated forms of J_H6 and J_H5 genes, respectively (Fig 1C). The deduced amino acid sequences of the D-J_H genes are depicted in Fig 1D, as segregated in CDR3 and FR4 stretches, according to Kabat et al.[49] The CDR3 sequences were highly divergent in length and composition. The HBL-2 and HBL-3 FR4 sequences were invariable in length and displayed two and one amino acid replacements, respectively.

The HBL-2 and HBL-3 MoAb Vλ and Jλ genes. Figure 2 depicts the nucleotide (A) and deduced amino acid (B) sequences of the HBL-2 and HBL-3 MoAb Vλ genes and those of the closest reported germline Vλ genes. The differences between sequences are summarized in Table 1. HBL-2 and HBL-3 MoAbs used two members of the Vλ1 subgroup, the Humiglv117 and Humiglv1L1 genes, respectively.[14,50] When compared with the germline gene, the HBL-2 Vλ1 gene sequence displayed nine and four nucleotide differences in the CDRs and FRs, respectively, yielding four and two amino acid replacements, and R:S mutation ratios of 1.2 and 1.0, respectively. When compared with the germline gene, the HBL-3 Vλ1 gene sequence displayed 39 nucleotide differences. These were scattered throughout the CDRs and FRs, yielding a total of 19 amino acid replacements and R:S mutation ratios of 1.1 and 0.8, respectively. Figure 2 depicts the nucleotide (C) and deduced amino acid (D) sequences of the MoAb Jλ segments and their respective germline Jλ templates. The HBL-2 MoAb used a Jλ2/Jλ3 segment in germline configuration; the HBL-3 MoAb used a Jλ1 segment with five nucleotide mutations resulting in three amino acid replacements.

Somatic mutations in the HBL-3 MoAb V_H segment. Because of conservation of the V_H4-21 gene in humans, the high number of nucleotide differences displayed by the HBL-3 V_H gene sequence when compared with that of the V_H4-21 germline gene, and the detection of mutations in the HBL-3 MoAb J_H5 and Jλ1 segments, we hypothesized that the HBL-3 MoAb V_H segment consisted of a somatically mutated form of the V_H4-21 gene. PCR amplifications were performed using ad hoc designed oligonucleotide primers and genomic DNA from the HBL-3 cell line or autologous fibroblasts. The sense V_H4-21 FR1 primer, encompassing an FR1 sequence (residues 10 to 27) shared by the V_H4-21 segment and the expressed HBL-3 V_H gene, was used in conjunction with the antisense V_H4-21 FR3 primer, encompassing an FR3 sequence (residues 250 to 269) shared by the germline V_H4-21 and the expressed HBL-3 genes. The two combined primers amplified DNA from both fibroblasts and HBL-3 cells. The molecular size of the amplified product (~260 bp) was consistent with that of the sequence spanning residues 10 to 269 of the V_H4-21 gene sequence (Fig 3A, lanes 1 and 2). The same antisense V_H4-21 FR3 primer was also used to amplify fibroblast DNA, in conjunction with the sense HBL-3 leader primer, encompassing a stretch of the leader sequence of HBL-3 V_H gene (residues −49 to −25) and differing in only one nucleotide from that of the corresponding area of the V58 gene, the V_HIV family member displaying the highest degree of identity with the V_H4-21 gene. The molecular size of the amplified product (~400 bp) was consistent with that of the sequence spanning residues −49 to 269 of the HBL-3 V_H gene, including the untranslated intervening intron (Fig 3A, lane 3).

The three DNA amplification products were analyzed for their ability to hybridize with the [γ-^{32}P]-labeled HBL-3 CDR1 oligonucleotide. This encompassed a stretch of the HBL-3 V_H gene FR1-CDR1 sequence that displayed seven putative mutations when compared with the corresponding germline V_H4-21 gene sequence. The [γ-^{32}P]-labeled HBL-3 CDR1 oligonucleotide strongly hybridized with the ~260 bp DNA amplified from the HBL-3 cell line (Fig 3B, lane 1), but not with the ~260 or ~400 bp DNA amplified from the autologous fibroblasts (Fig 3B, lanes 2 and 3, respectively). To identify the autologous germline V_H gene that putatively gave rise to the expressed HBL-3 V_H gene, the product amplified from fibroblast DNA using the sense V_H4-21 FR1 and the antisense V_H4-21 FR3 primers was cloned

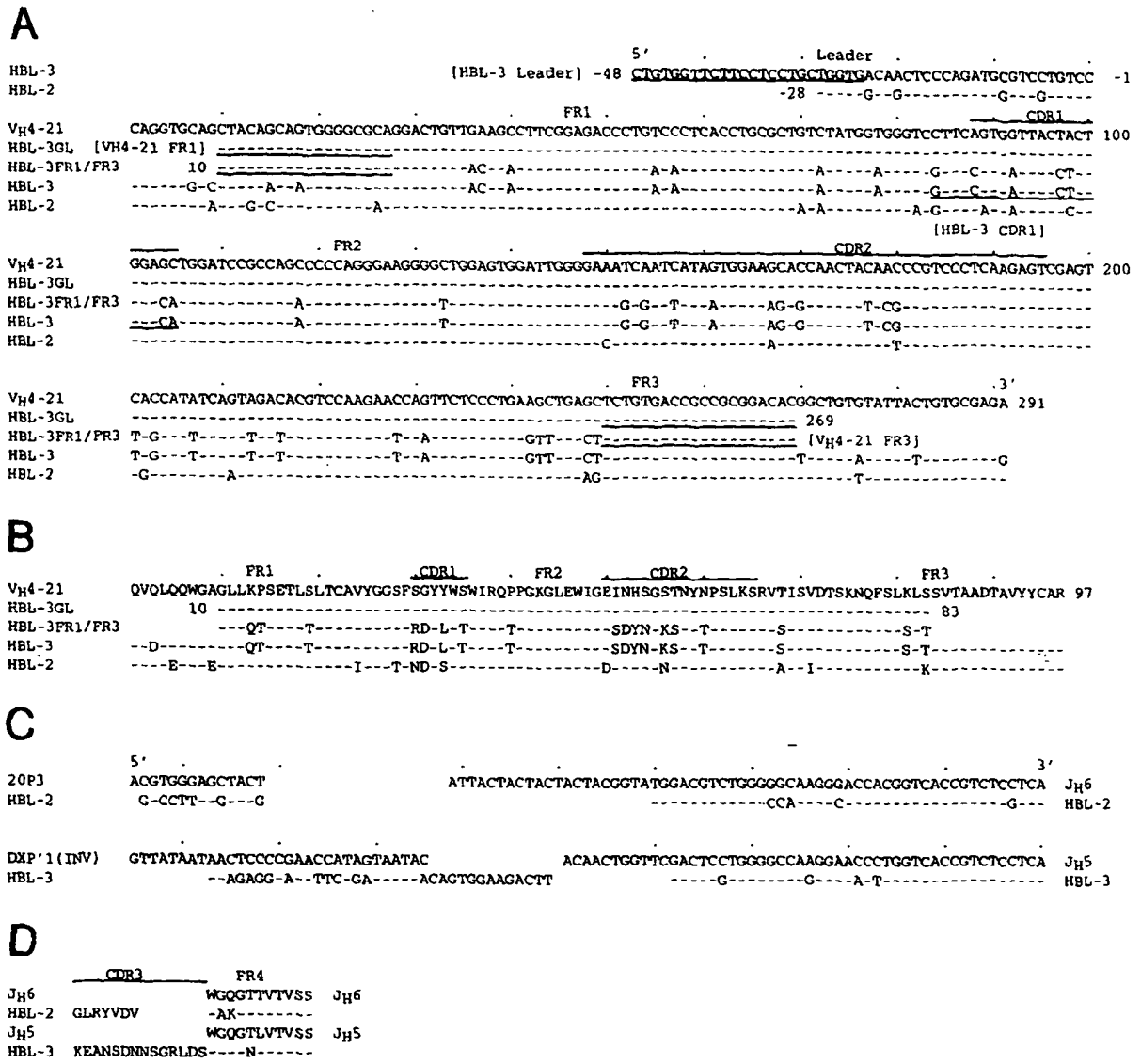


Fig 1. Top clusters: nucleotide (A) and deduced amino acid (B) sequences of the V_H genes used by the HBL-2 and HBL-3 IgM MoAbs. The top sequence is given for comparison and represents the published germline V_H gene (V_H4-21) displaying the highest degree of identity to the expressed V_H genes. The V_H4-21 gene belongs to the V_HIV family. Dashes indicate identity. Solid lines on the top of each cluster depict CDR. Underlined nucleotides define the sequences or the reverse complement sequences of the primers adopted for PCR amplification. HBL-3 GL and HBL-3 FR1/FR3 are the sequences amplified from fibroblast and HBL-3 cell genomic DNA, respectively, using the combination of the V_H4-21 FR1 and the V_H4-21 FR3 oligonucleotide primers (see Materials and Methods for details). Bottom clusters: nucleotide (C) and deduced amino acid (D) sequences of the D and J_H genes used by the HBL-2 and HBL-3 IgM MoAbs. Germline D genes are given for comparison. Dashes indicate identity. DXP'1 (Inv) is the reverse complement of the germline DXP'1 sequence. The present sequences are available from EMBL/GenBank/DDBJ under accession numbers L29115 and L29116.

and sequenced. The sequences of the eight independent clones were all identical to each other and to that of the V_H4-21 germline gene throughout the overlapping area (residues 28 through 249) (Fig 1A and B; HBL-3GL sequence). DNA amplified from the HBL-3 cell line using the sense V_H4-21 FR1 and the antisense V_H4-21 FR3 primers was also cloned and sequenced. The sequences of three independent clones were identical to each other and to that of the expressed HBL-3 V_H gene throughout the overlapping area (residues 28 through 249) (Fig 1A and B; HBL-3 FR1/FR3 sequence). These experiments proved that the expressed HBL-3 V_H segment was somatically point-mutated, and suggested that V_H4-21 is the germline gene that gave rise to it.

Ag-selection of the IgV genes expressed by the HBL-2 and HBL-3 BL cells. In absence of negative or positive selective pressure on a gene product, R and S mutations are scattered randomly throughout the protein sequence. If a DNA segment displays a number of R mutations lower than

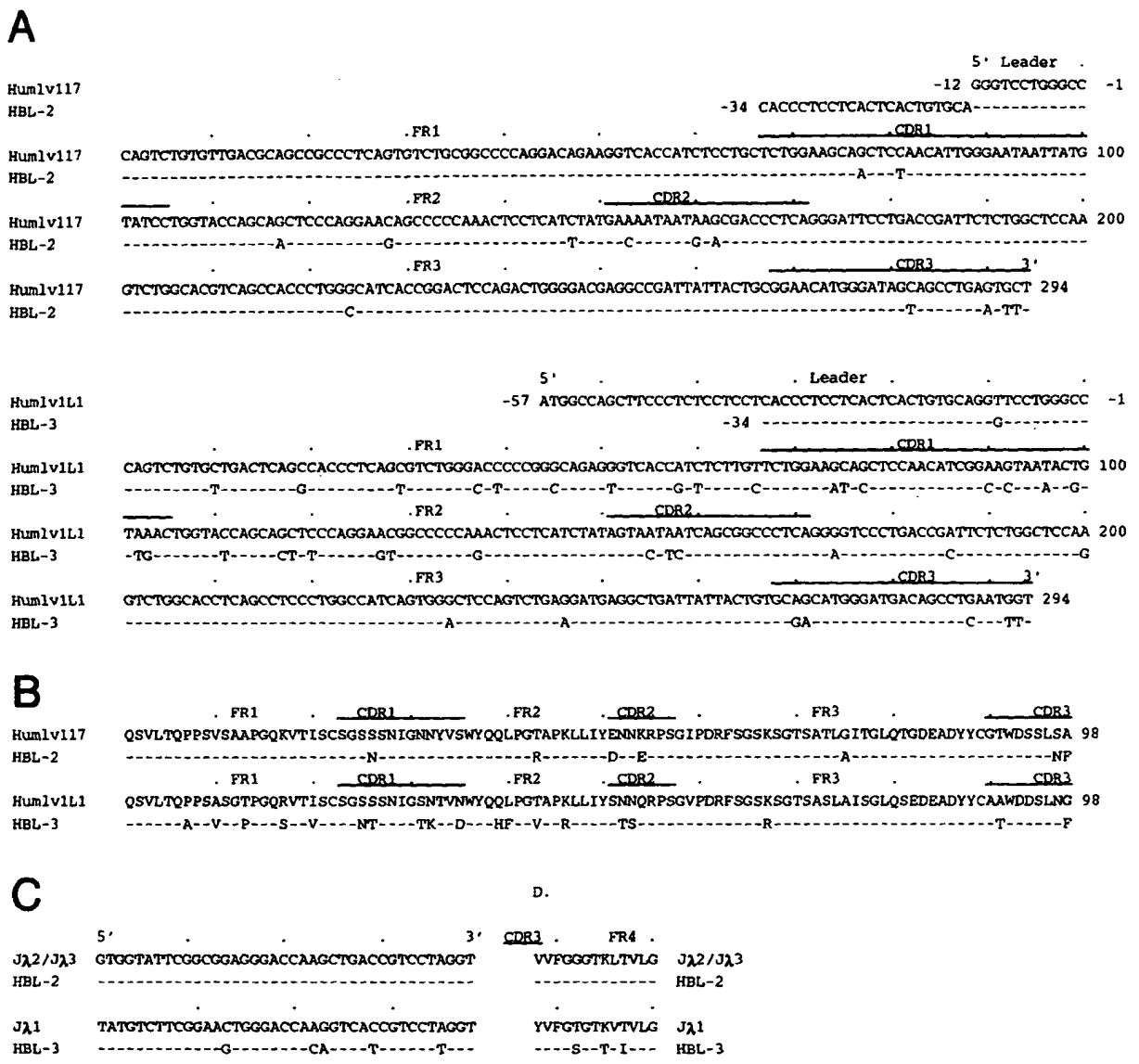


Fig 2. Top clusters: nucleotide (A) and deduced amino acid (B) sequences of the Vλ genes used by the HBL-2 and HBL-3 IgM MoAbs. In each cluster, the top sequence is given for comparison and represents the published germline Vλ gene displaying the highest degree of identity with the expressed Vλ genes. Dashes indicate identity. Solid lines on the top of each cluster depict CDR. The Humlv117 and Humlv1L1 genes belong to the Vλ1 subgroup. Bottom clusters: nucleotide (C) and deduced amino acid (D) sequences of the Jλ segments used by the HBL-2 and HBL-3 MoAbs. Dashes indicate identity. The present sequences are available from EMBL/GenBank/DDBJ under accession numbers L29113 and L29114.

that expected by chance only, it is likely that pressure to maintain the germline-encoded protein structure was exerted. Conversely, if a DNA segment displays a number of R mutations higher than that expected by chance only, it is likely that a positive pressure to select R mutations was exerted. The numbers of expected R mutations in the HBL-2 and HBL-3 MoAb V_H and Vλ segment CDRs and FRs were calculated using the formula $n \times Rf \times CDRf$ (or FRf), where n is the total number of observed mutations, Rf is the expected proportion of R mutations (0.75),[51] and $CDRf$ is the relative size of the CDRs (or FRs) (0.23 and 0.77 for V_H4-21 CDRs and FRs, respectively; 0.29 and 0.71 for Vλ1 CDRs and FRs, respectively). Both HBL-2 and HBL-3 MoAb V segments displayed higher and lower numbers of R mutations in the CDRs and FRs, respectively, than those theoretically expected (Table 1). Because of the primary role played by the V_H4.21 segment in the binding to the i/I Ag, we calculated the probabilities that the excess and the scarcity of R mutations arose by chance in the HBL-2 and HBL-3 MoAb V_H segment CDRs and FRs, respectively, using the binomial distribution model $P = [n!/k!(n - k)!]\, q^k\, (1 - q)^{n-k}$, where q is the probability an R mutation will locate

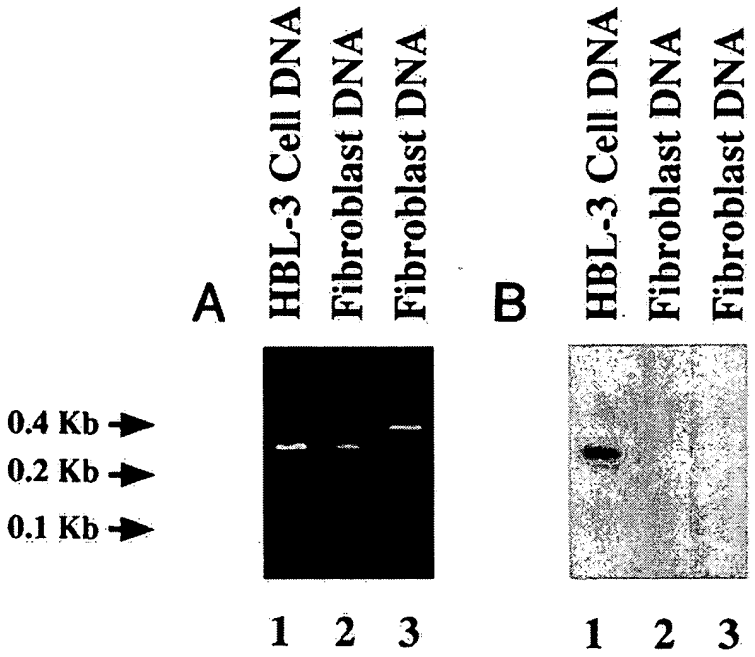


Fig 3. Evidence for somatic mutations in the HBL-3 V_H gene. (A) Ethidium bromide staining of amplified DNA after fractionation in agarose gel electrophoresis. Using the V_H4-21 FR1 and V_H4-21 FR3 oligonucleotide primers, amplification products of identical and appropriate size were obtained by priming genomic DNA from the HBL-3 cells (lane 1) and autologous fibroblasts that had been established in primary culture from the BL bioptic sample of the patient whose neoplastic B cells gave rise to the HBL-3 cell line (lane 2). Using the HBL-3 leader and V_H4-21 FR3 oligonucleotide primers, a product of appropriate size was amplified by priming genomic DNA from fibroblasts (lane 3). (B) Southern blot hybridization of the PCR products shown in (A) with ^{32}P-labeled oligonucleotide HBL-3 CDR1 probe encompassing an FR1-CDR1 sequence of the HBL-3 MoAb V_H gene. A strong positive hybridization signal was detected only with DNA amplified from the HBL-3 MoAb-producing B-cell line (lane 1), but not with DNA amplified from autologous fibroblasts (lanes 2 and 3).

to the CDRs ($q = 0.23 \times 0.75$) or FRs ($q = 0.77 \times 0.75$), and k is the number of observed R mutations in the CDRs or FRs.[52] The likelihood that the excess R mutations arose by chance in the V_H segment CDR were $P = .11$ in HBL-2 MoAb and $P = .06$ in HBL-3 MoAb. The probability that the scarcity of R mutations in the V_H segment FR resulted from chance were $P = .03$ in HBL-2 MoAb and $P = .00000002$ in HBL-3 MoAb. In their original report on the application of the binomial distribution model to the analysis of the R point-mutations in Ag-selected IgV segment CDRs, Shlomchik et al[52] suggested that the observed number of FR R mutations should be doubled to account for the fact that some of these mutational events will never be observed because they are deleterious to the Ig structure. Although this correction was inferred from some experimental observations, it is approximate and may not be applicable as such to all Ig V_H genes. When the analysis of the V_H segment somatic mutation pattern was performed with the adjustment of doubling the number of observed FR R mutations, the probabilities that excess R mutations arose by chance in the V_H segment CDR were $P = .49$ and $P = .35$ in HBL-2 MoAb and HBL-3 MoAb, respectively.

DISCUSSION

In the present studies, we established from bioptic specimens of 2 AIDS patients with BL two MoAb-producing cell lines representative of the respective tumors, and analyzed the Ag-binding activity and the V segment structure of these MoAbs. We found that both IgM MoAbs were cold agglutinins highly specific for the i blood group determinant, and both MoAbs bore Ag-combining sites consisting of point-mutated V_H4-21 segments in conjunction with Vλ1 segments.

The exquisite specificity of the HBL-2 and HBL-3 IgM MoAb cold agglutinins for the i repetitive N-acetyllactosamine units was strengthened by the MoAb failure to bind any of the other eight self Ags and nine foreign Ags tested. The putative use of the V_H4-21 gene segment by the HBL-2 and HBL-3 MoAbs is consistent with the use of the same segment by the majority of the reported cold agglutinins from patients with idiopathic cold agglutinin disease, FL, and Waldenstrom's macroglobulinemia.[1,2,53-56] A primary role of the V_H4-21 segment in the binding to the i Ag is further supported by the divergence in composition and length of the H chain CDR3 sequences as well as by the heterogeneity of the V_L segments of the present two IgM and the 10 reported V_H4-21$^+$ cold agglutinins, which use a V_κI segment three times, a V_κII segment once, a V_κIIIa segment once, a V_κIIIb four times, and a Vλ1 segment once.[1,54-56] Thus, the V_H4-21 gene restriction in the cold agglutinin system may result from a selection process based on an inherent affinity of this V_H gene product for the i/I carbohydrate structure.[1,2,54,56] Nevertheless, the V_H4-21 segment is not an absolute requirement for i/I-dependent RBC agglutination, because cold agglutinins using V_H segments of the V_HIII family have been reported.[57] The restricted usage of the V_H4-21 by anti-i/I cold agglutinins is intriguing and may be related to the overrepresentation of the V_H4-21–expressing clones in the normal B-cell repertoire, as determined by the V_H4-21–related 9G4 idiotype studies in the circulating blood of adults, as well as cord blood and fetal tissues.[4] In this regard, it is not known whether these circulating V_H4-21–expressing B cells also have anti-i/I specificity. If this were the case, one could speculate that the abundant representation of V_H4-21–expressing B cells in the periphery results from positive selection by i/I Ag, which are present not only on RBCs but

also on lymphocytes.[58] Expression of the i Ag on human erythrocytes is developmentally regulated. It is maximal in fetal and neonatal life and decreases in adult life, in which the expression of the I Ag prevails.[59]

An Ag-driven process of clonal selection may play a role in the emergence and/or expansion of certain neoplastic B lymphocytes. Consistent with their putative germinal center origin, FL B cells, which represent the neoplastic equivalents of the elements recruited in a secondary Ag-specific response, display somatic mutations that resemble in nature and distribution those characteristic of an affinity maturation process.[22-24] A recent thorough documentation of the somatic mutation and clonal evolution of an FL expressing a V_H4-21 gene, the antigenic specificity of which had not been determined, showed at least three amino acid mutations, the anti-i/I "characteristic" Gly to Asn mutation at position 31, a Val to Ile mutation at position 71, and a Ser to Thr mutation at position 83, which are shared by the HBL-2 and HBL-3 MoAb V_H4-21 segments.[24] A role for clonal selection by self Ag during the evolution of anti-"Pr_2"-specific B-cell lymphoma has been documented in detail.[5,6] The Pr_2 Ag is a sialoglycoprotein and provides, along with the multiple N-acetyllactosamine i/I Ag, the target for autoimmune phenomena that occur in association with several human clonal B-cell disorders.

The sequences of the DNA amplified from the HBL-3 MoAb-producing cell line and autologous fibroblast genomic DNA, using the HBL-3 leader, V_H4-21 FR1, and V_H4-21 FR3 primers, as well as the differential hybridization of the HBL-3 CDR1 oligonucleotide (encompassing an FR1-CDR1 sequence of the HBL-3 MoAb V_H segment) with the above amplification products, formally proved the mutated status of the HBL-3 V_H segment, and suggested that V_H4-21 was the germline gene that gave rise to it. The somatically mutated status of the HBL-3 and, possibly, HBL-2 MoAbs was further strengthened by the high degree of conservation of the V_H4-21 gene sequence in humans, and the extension of the point-mutations to the, in general, highly conserved, J_H and/or Jλ segments. An Ag-selection of the point-mutations in the HBL-2 and HBL-3 MoAb V_H segments was suggested by the differential R:S mutation ratios in the CDR and FR (HBL-2 MoAb, 5.0 and 1.1, respectively; HBL-3 MoAb, 2.2 and 0.3, respectively) and the accumulation in the CDR of the HBL-2 and/or HBL-3 V_H segments of amino acid replacements that are shared by other anti-i/I cold agglutinins and might have increased the affinity of the V_H4-21 segment for the i Ag, including the Gly to Asp mutation at position 31, which is shared by the FS-1, FS-2, FS-4, and KAU V_H cold agglutinins,[54,55] the Ser to Thr mutation at position 35, which is shared by FS-1, FS-4, and FS-6 cold agglutinins, and the His to Tyr at position 53, which is shared by the FS-3 cold agglutinin.[55] However, a positive clonal selection of R mutations in the HBL-2 and HBL-3 MoAb V_H segment CDRs was not further substantiated by the statistical analysis according to the binomial distribution model with the correction for FR R mutations, as proposed by Shlomchik et al.[52] This finding may be consistent with a putatively inherent anti-i/I activity of the unmutated V_H4-21 gene product, and, perhaps, a clonal selection against R mutations in the HBL-2 and HBL-3 MoAb V_H segments CDRs, similar to that shown for other B-cell tumor anti-RBC autoantibodies.[6] The substitution of the Val with an Ile in the HBL-3 MoAb V_H segment FR1 Ala-Val-Tyr (residues 23 to 25) triplet, which provides the structural correlates for the anti-idiotypic 9G4 antibody binding, as recently shown by Potter et al,[60] is possibly responsible for the lack of 9G4 reactivity of the HBL-2 MoAb.

In the present AIDS-associated BLs, it is unclear whether the initiation of the anti-i/I Ag autoantibody response constituted a crucial event in the neoplastic transformation. The putative anti-self Ag clonal expansion and selection may have preceded the genetic accident, ie, c-*myc* proto-oncogene chromosomal translocation.[61,62] Alternatively, in these BLs, the specific B-cell expansion and selection may have followed the chromosomal translocation, resembling the series of events that have been paradigmatically illustrated in relationship to *bcl*-2 proto-oncogene chromosomal t(14;18) translocation by Zelenetz et al[23] in a FL for which, however, a specific Ag could not be identified. Knowledge of the sequential order of activating, proliferating, and transforming events, including c-*myc* translocation and activation, Ag-dependent B-cell amplification, somatic hypermutation, and clonal selection, is crucial for a better understanding of the molecular pathogenesis of AIDS BL and, possibly, other BLs. These issues could be best addressed by the use of a tumor-specific Ig H chain CDR3 sequence oligonucleotide to identify tumor-related Ig V_H-D-J_H sequences in nonmalignant B-cell progenitors.

ACKNOWLEDGMENT

We thank Dr M. Gorny for testing the HBL-2 and HBL-3 MoAbs for binding to HIV-1, CMV, and parvovirus.

REFERENCES

1. Silberstein LE, Jefferies LC, Goldman J, Friedman D, Moore JS, Nowell PC, Roelcke D, Pruzanski W, Roudier J, Silverman GJ: Variable region gene analysis of pathologic human autoantibodies to the related i and I red blood cell antigens. Blood 78:2372, 1991

2. Pascual V, Victor K, Lelsz D, Spellberg MB, Hamblin TJ, Thompson KM, Randen I, Natvig J, Capra JD, Stevenson FK: Nucleotide sequence analysis of the V regions of two IgM cold agglutinins. Evidence that the V_H4-21 gene segment is responsible for the major cross-reactive idiotype. J Immunol 146:4385, 1991

3. Miller DG: The association of immune disease and malignant lymphoma. Ann Intern Med 66:507, 1967

4. Stevenson FK, Smith GJ, North J, Hamblin TJ, Glennie MJ: Identification of normal B-cell counterparts of neoplastic cells which secrete cold agglutinins of anti-I anti-i specificity. Br J Haematol 72:9, 1989

5. Silberstein LE, Litwin S, Carmack CE: Relationship of variable region genes expressed by a human B cell lymphoma secreting pathologic anti-PR_2 erythrocyte autoantibodies. Exp Med 169:1631, 1989

6. Friedman DF, Cho EA, Goldman J, Carmack CE, Besa EC, Hardy RR, Silberstein LE: The role of clonal selection in the pathogenesis of an autoreactive human B cell lymphoma. J Exp Med 174:525, 1991

7. Kaye BR: Rheumatologic manifestations of infection with human immunodeficiency virus (HIV). Ann Intern Med 111:158, 1989

8. Freter CE: Acquired immunodeficiency syndrome-associated lymphomas. J Natl Cancer Inst Monogr 10:45, 1990

9. Raphael BG, Knowles DM: Acquired immunodeficiency syndrome-associated non-Hodgkin's lymphoma. Semin Oncol 17:361, 1990

10. Shlomchik MJ, Marshak-Rothstein A, Wolfowicz CB, Rothstein TL, Weigert MG: The role of clonal selection and somatic mutation in autoimmunity. Nature 328:805, 1987

11. Shlomchik M, Mascelli M, Shan H, Radic MZ, Pisetsky D, Marshak-Rothstein A, Weigert MG: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic mutation. J Exp Med 171:265, 1990

12. Harindranath N, Goldfarb IS, Ikematsu H, Burastero SE, Wilder RL, Notkins AL, Casali P: Complete sequence of the genes encoding the V_H and V_L regions of low and high affinity monoclonal IgM and IgA1 rheumatoid factors produced by $CD5^+$ B cells from a rheumatoid arthritis patient. Int Immunol 3:865, 1991

13. Randen I, Brown D, Thompson KM, Hughens-Jones N, Pascual V, Victor K, Capra JD, Forre O, Natvig JB: Clonally related IgM rheumatoid factors, undergo affinity maturation in the rheumatoid synovial tissue. J Immunol 148:3296, 1992

14. Olee T, Liu EW, Huang D-F, Soto-Gil RW, Deftos M, Kozin F, Carson DA, Chen PP: Genetic analysis of self-associating immunoglobulin G rheumatoid factors from two rheumatoid synovia implicates an antigen driven response. J Exp Med 167:840, 1992

15. van Es JH, Aanstoot H, Gmelig-Meyling FHJ, Derksen RHWM, Logtenberg T: A human systemic lupus erythematosus-related anti-cardiolipin/single-strand DNA autoantibody is encoded by somatically mutated variant of the developmentally restricted 51P1 V_H gene. J Immunol 149:2234, 1992

16. Mantovani L, Wilder RL, Casali P: Human rheumatoid B-1a ($CD5^+$ B) cells make somatically hypermutated high affinity IgM rheumatoid factors. J Immunol 151:473, 1993

17. McKean D, Huppi K, Bell M, Staud L, Gerhard W, Weigert MG: Generation of antibody diversity in the immune response of Balb/c mice to influenza virus hemagglutinin. Proc Natl Acad Sci USA 81:3180, 1984

18. Berek C, Milstein C: Mutation drift and repertoire shift in the maturation of the immune response. Immunol Rev 96:23, 1987

19. Manser T: Evolution of antibody structure during the immune response. The differentiative potential of a single B lymphocyte. J Exp Med 170:1211, 1989

20. Ueki Y, Goldfarb IS, Harindranath N, Gore M, Koprowski H, Notkins AL, Casali P: Clonal analysis of a human antibody response. Quantitation of precursors of antibody-producing cells and generation and characterization of monoclonal IgM, IgG, and IgA to rabies virus. J Exp Med 171:19, 1990

21. Ikematsu H, Harindranath N, Ueki Y, Notkins AL, Casali P: Clonal analysis of a human antibody response. II. Sequences of the V_H genes of human monoclonal IgM, IgG and IgA to rabies virus reveal preferential utilization of the V_HIII family members and somatic hypermutation. J Immunol 150:1325, 1993

22. Levy S, Mendel E, Kon S, Avnur Z, Levy R: Mutational hot spots in Ig V region genes of human follicular lymphomas. J Exp Med 168:475, 1988

23. Zelenetz AD, Chen TT, Levy R: Clonal expansion in follicular lymphoma occurs subsequent to antigenic selection. J Exp Med 176:1137, 1992

24. Bahler DW, Levy R: Clonal evolution of a follicular lymphoma: Evidence for antigen selection. Proc Natl Acad Sci USA 89:6770, 1992

25. Pelicci P-G, Knowles DM, Arlin ZA, Wieczorek R, Luciw P, Dina D, Basilico C, Dalla-Favera R: Multiple monoclonal B cell expansions and c-*myc* oncogene rearrangements in acquired immune deficiency syndrome-related lymphoproliferative disorders. J Exp Med 164:2049, 1986

26. Gaidano G, Dalla-Favera R: Biologic aspects of human immunodeficiency virus-related lymphoma. Curr Opin Oncol 4:900, 1992

27. Knowles DM, Chadburn A: Lymphadenopathy and the lymphoid neoplasms associated with the acquired immune deficiency syndrome (AIDS), in Knowles DM (ed): Neoplastic Hematology. Baltimore, MD, Williams & Wilkins, 1992, p 773

28. Gaidano G, Parsa NZ, Tassi V, Della-Latta P, Chaganti RSK, Knowles DM, Dalla-Favera R: In vitro establishment of AIDS-related lymphoma cell lines: Phenotypic characterization, oncogene and tumor suppressor gene lesions, and heterogeneity in Epstein-Barr virus infection. Leukemia 7:1621, 1993

29. Korsmeyer SJ, Hyeter PA, Ravetch JV, Poplack DG, Waldmann TA, Leder P: Developmental hierarchy of immunoglobulin gene rearrangement in human leukemic pre-B cells. Proc Natl Acad Sci USA 78:7096, 1981

30. Dalla-Favera R, Bregni M, Erikson J, Patterson D, Gallo RC, Croce CM: Human *c-myc* oncogene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proc Natl Acad Sci USA 79:7824, 1982

31. Gorny MK, Gianakakos V, Sharpe S, Zolla-Pazner S: Generation of human monoclonal antibodies to human immunodeficiency virus. Proc Natl Acad Sci USA 86:1624, 1989

32. Nakamura M, Burastero SE, Notkins AL, Casali P: Human monoclonal rheumatoid factor-like antibodies from CD5 (Leu-1)$^+$ B cells are polyreactive. J Immunol 140:4180, 1988

33. Kasaian MT, Ikematsu H, Casali P: Identification and analysis of a novel CD5− B lymphocyte subset producing natural antibodies. J Immunol 148:2690, 1992

34. Nakamura M, Burastero SE, Ueki Y, Larrick JW, Notkins AL, Casali P: Probing the normal and autoimmune B cell repertoire with Epstein-Barr virus. Frequency of B cells producing monoreactive high affinity autoantibodies in patients with Hashimoto's disease and systemic lupus erythematosus. J Immunol 141:4165, 1988

35. Silberstein LE, Jefferies LC, Goldman J, Spitalnik SL: Production of carbohydrate-specific human monoclonal antibodies in vitro. Methods Enzymol 179:299, 1989

36. Stevenson FK, Wrightham M, Glennie MJ, Jones DB, Cattan AR, Feizi T, Spelleberger MB, Hamblin TJ, Stevenson GT: Antibodies to shared idiotypes as reagents for analysis and therapy for human B cell tumors. Blood 68:430, 1986

37. Parson WR: Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 183:63, 1990

38. Sanz I, Kelly P, Williams C, Scholl S, Tucker P, Capra JD: The smaller human V_H gene families display remarkably little polymorphism. EMBO J 8:3741, 1989

39. Lee KH, Matsuda F, Kinashi T, Kodaira M, Honjo T: A novel family of variable region genes of the human immunoglobulin heavy chain. J Mol Biol 195:761, 1987

40. Magrath IT: The pathogenesis of Burkitt's Lymphoma. Adv Cancer Res 55:133, 1990

41. van Es JH, Heutink M, Aanstoot H, Logtenberg T: Sequence analysis of members of the human Ig V_H4 gene family derived from a single V_H locus. Identification of novel germ-line members. J Immunol 149:492, 1992

42. Ravetch JV, Siebenlist U, Korsmeyer SJ, Waldmann T, Leder P: Structure of the human immunoglobulin μ locus: Characterization of embryonic and rearranged J and D genes. Cell 27:583, 1981

43. Siebenlist V, Ravetch JV, Korsmeyer SJ, Waldmann T, Leder P: Human immunoglobulin D segments encoded in tandem mutigene families. Nature 294:631, 1981

44. Ichihara Y, Matsuoka H, Kurosawa Y: Organization of human immunoglobulin heavy chain diversity gene loci. EMBO J 7:4141, 1988

45. Buluwela L, Albertson DG, Sherrington P, Rabbitts PH, Spurr N, Rabbitts TH: The use of chromosomal translocations to study

human immunoglobulin gene organization: Mapping D_H segments within 35 kb of the Cμ gene and identification of a new D_H locus. EMBO J 7:2003, 1988

46. Matsuda F, See KH, Nakai S, Sato T, Kodaira M, Zong SQ, Ohno H, Fukuhara S, Honjo T: Dispersed localization of D segments in the human immunoglobulin heavy-chain locus. EMBO J 7:1047, 1988

47. Schroeder HW Jr, Hillson JL, Perlmutter RM: Early restriction of the human antibody repertoire. Science 238:791, 1987

48. Meek KD, Hasemann CA, Capra JD: Novel rearrangements at the immunoglobulin D locus. Inversion and fusion add to IgH somatic diversity. J Exp Med 170:39, 1989

49. Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C: Sequences of Proteins of Immunological Interest. Washington, DC, US Department of Health and Human Services, 1991

50. Siminovitch KA, Misener V, Kwong PC, Song Q-L, Chen PP: A natural antibody is encoded by germline heavy and lambda light chain variable region genes without somatic mutation. J Clin Invest 84:1675, 1989

51. Jukes TH, King JL: Evolutionary nucleotide replacements in DNA. Nature 281:605, 1979

52. Shlomchik MJ, Aucoin AH, Pisetsky DS, Weigert MG: Structure and function of anti-DNA autoantibodies derived from a single autoimmune mouse. Proc Natl Acad Sci USA 84:9150, 1987

53. Silverman GJ, Carson DA: Structural characterization of human monoclonal cold agglutinins: Evidence for a distinct primary sequence-defined V_H4-21 idiotype. Eur J Immunol 20:351, 1990

54. Leoni J, Ghiso J, Goni F, Frangione B: The primary structure of the Fab fragment of protein KAU, a monoclonal immunoglobulin M cold agglutinin. J Biol Chem 266:2836, 1991

55. Pascual V, Victor K, Spellerberg M, Hamblin TJ, Stevenson FK, Capra JD: V_H restriction among human cold agglutinins. The V_H4-21 gene segment is required to encode anti-I and anti-i specificities. J Immunol 149:2337, 1992

56. Pascual V, Capra JD: V_H4-21, a human V_H gene segment overrepresented in the autoimmune repertoire. Arthritis Rheum 35:11, 1992

57. Jeffreys LC, Carachidi CM, Goldman J, Silberstein LE, Naturally occurring anti-*i*/*I* cold agglutinins may be encoded by different V_HIII genes as well as by V_H4.21 gene segments. J Clin Invest 92:2821, 1993

58. Grillot-Courvalin C, Brouet J-C, Piller F, Rassenti LZ, Labaume S, Silverman GJ, Silberstein L, Kipps TJ: An anti-B cell autoantibody from Wiskott-Aldrich syndrome which recognizes i blood group specificity on normal human B cells. Eur J Immunol 22:1781, 1992

59. Fukuda M, Fukuda MN, Papayannopoupou T, Hakomori SI: Membrane differentiation in human erythroid cells: Unique profiles of cell surface glycoproteins expressed in erythroblasts in vitro from three ontogenic stages. Proc Natl Acad Sci USA 77:3473, 1980

60. Potter KN, Li Y, Pascual V, Williams RC Jr, Byres LC, Spellberg M, Stevenson FK, Capra JD: Molecular characterization of a cross-reactive idiotope on human immunoglobulins utilizing the V_H4-21 gene segment. J Exp Med 178:1419, 1993

61. Pelicci P-G, Knowles DM, Magrath IT, Dalla-Favera R: Chromosomal breakpoints and structural alterations of the *c-myc* locus differ in endemic and sporadic forms of Burkitt's lymphoma. Proc Natl Acad Sci USA 83:2984, 1986

62. Neri A, Barriga F, Knowles DM, Magrath IT, Dalla-Favera R: Different regions of the immunoglobulin heavy-chain locus are involved in chromosomal translocations in distinct pathogenetic forms of Burkitt's lymphoma. Proc Natl Acad Sci USA 85:2748, 1988

Exhibit L

Pergamon

Molecular Immunology, Vol. 32, No. 14/15, pp. 1105–1122, 1995
Elsevier Science Ltd
Printed in Great Britain
0161-5890/95 $9.50 + 0.00

0161-5890(95)00071-2

THE HUMAN ANTIBODY REPERTOIRE: HEAVY AND LIGHT CHAIN VARIABLE REGION GENE USAGE IN SIX ALLOANTIBODIES SPECIFIC FOR HUMAN HLA CLASS I AND CLASS II ALLOANTIGENS

JENNIFER S. ANDRIS,*† SHEENA R. ABRAHAM,* VIRGINIA PASCUAL,* MARIA P. PISTILLO,‡ STEFANO MANTERO,‡ GIOVANNI B. FERRARA‡ and J. DONALD CAPRA*§

*Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.; †Graduate Program in Immunology, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.; ‡Centro di Biotecnologie Avanzate, Istituto Naztionale per la Ricerca sul Cancro, Genova, Italy

(*First received* 24 *February* 1995; *accepted in revised form* 5 *May* 1995)

Abstract—Peripheral blood B lymphocytes have been isolated from healthy individuals who were immunized with lymphocytes from HLA-incompatible donors and transformed with Epstein–Barr virus to produce human monoclonal cell lines specific for human HLA molecules. The cell lines have been previously characterized and are known to bind to various class I and class II alloantigens. In this report we describe the molecular characterization of the heavy and light chain variable region gene segments that are utilized by these monoclonal antibodies. Using the polymerase chain reaction and primer pairs specific for the respective constant region and V_H or V_L family, rearranged variable region gene segments were amplified from cDNA from individual cell lines. Products were then subcloned, sequenced and analysed for gene usage and apparent somatic mutation. The results show that the V_H 3 gene family predominates in a group of six heavy chains (four out of six) with one V_H1 and one V_H4 gene segment. The light chain variable region gene family usage is more diverse with 2 V_k3, 1 V_k1, 2 V_λ2 and 1 V_λ3. The extent of apparent somatic mutation is minimal, relative to our previous observations in a group of high affinity human monoclonal antibodies specific for pathogenic organisms.

Key words: HLA, alloantigen, alloantibody, variable region, J region, D segment, human antibody repertoire, human monoclonal antibody, apparent somatic mutation.

INTRODUCTION

One of the unique and critical features of the humoral immune response is the innate ability to construct an immunoglobulin molecule with virtually any specificity. The formation of an antibody response begins with the germline rearrangement of variable $(V_H)^3$, diversity (D), and joining (J_H) gene segments for the heavy (H) chain and variable (V_L) and joining (J_L) gene segments for the light (L) chain and ends with the antigen-driven selection of a B lymphocyte which possesses a unique binding specificity. In the interim the association of the heavy and light chains and the initiation of a number of mechanisms results in a repertoire of antibody-expressing B cells with an essentially infinite number of specificities (Tonegawa, 1983). The gene segments which comprise a complete immunoglobulin molecule are found within three separate gene complexes—heavy chain, κ light chain and λ light chain—all of which are located on distinct chromosomes and consist of multiple germline gene segments. These gene segments are grouped into gene families based on nucleotide sequence homology. There are seven V_H (reviewed in Pascual and Capra, 1990; van Dijk *et al.*, 1993); seven V_κ (Zachau, 1989); and at least 10 V_λ (Anderson *et al.*, 1984; Williams and Winter, 1993; Stiernholm *et al.*, 1994; Chuchana *et al.*, 1990) gene families. A considerable amount of knowledge exists regarding the rearrangement process and the subsequent differentiation of a B lymphocyte, but one predominant question remains unanswered: why do humans maintain 50–100 distinct variable region germline gene segments for each chain?

In order to address this question a number of human immunoglobulin repertoire studies involving the structural analysis of the H and L chain variable region gene segments have been reported. These include myeloma proteins (Capra and Kehoe, 1975), fetal rearrangements (Schroeder *et al.*, 1987; Cuisinier *et al.*, 1989; Schroeder and Wang, 1990; Pascual *et al.*, 1993), cord blood

§Author to whom correspondence should be addressed at: University of Texas Southwestern Medical Center, Department of Microbiology, 5323 Harry Hines Blvd, Dallas, TX 75235-9048, U.S.A.

rearrangements (Mortari *et al.*, 1992, 1993), hybridomas and Epstein–Barr virus (EBV)-transformed B cells and combinatorial antibody libraries producing a wide variety of autoantibodies (reviewed in Dersimonian *et al.*, 1990); alloantibodies (Hughes-Jones *et al.*, 1991; Thompson *et al.*, 1991; Larrick *et al.*, 1989*b*); and antibodies which bind to a number of different exogenous antigens, i.e. pathogenic microorganisms (Newkirk *et al.*, 1988; Larrick *et al.*, 1989*a*; Felgenhauer *et al.*, 1990; Andris *et al.*, 1991, 1992, 1993; Burton *et al.*, 1991; Marasco *et al.*, 1990; Barbas *et al.*, 1992, 1993; Zebedee *et al.*, 1992; Ikematsu *et al.*, 1993; Barrett *et al.*, 1992; Schroeder *et al.*, 1992; Insel *et al.*, 1992; Moran *et al.*, 1993; Scott *et al.*, 1992; Gillies *et al.*, 1989; Persson *et al.*, 1991; Ayala-Avila *et al.*, 1993; Bogard *et al.*, 1993; Rioux *et al.*, 1994; Lewis *et al.*, 1993). Some circumstances suggest that there is some bias in gene expression, i.e. expression of different gene families does not reflect the size of the family, nor are functional members of any given family expressed equally (Pascual and Capra, 1992). Such gene segment bias has been well documented in the murine system, as demonstrated by fetal rearrangements (Yancoupoulos *et al.*, 1984; Perlmutter *et al.*, 1985), and antibodies specific for antigens such as dextran and galactan (Schilling *et al.*, 1980; Rudikoff *et al.*, 1983). In human antibodies this bias has been seen predominantly in fetal rearrangements (Schroeder *et al.*, 1987; Cuisinier *et al.*, 1989; Schroeder and Wang, 1990; Pascual *et al.*, 1993), autoantibodies (reviewed in Dersimonian *et al.*, 1990; Pascual and Capra, 1992) and *Haemophilus influenzae* type b capsular polysaccharide antibodies (reviewed in Insel *et al.*, 1992; Scott *et al.*, 1992). Due to the prevalence of V region cross-reactive idiotypes (CRI) in several different autoimmune diseases and the study of V segment gene usage in human monoclonal antibodies, it has been hypothesized that multiple gene segments exist for the generation of distinct repertoires of B cells. However, studies involving the structural analysis of monoclonal antibodies to viral and bacterial antigens would suggest that many of the same variable region gene segments utilized in the aforementioned repertoires are also recruited in response to exogenous antigens. On the contrary, some human V_H gene segments seem to be over-represented in the autoimmune repertoire. The V4-34 (V_H4–21) gene segment, for example, has only been found in autoantibodies, particularly cold agglutinins (Pascual *et al.*, 1991, 1992*b*), anti-DNA antibodies (van Es *et al.*, 1991), RF (Silberstein *et al.*, 1991; Pascual *et al.*, 1992*a*) and human red blood cell specific alloantibodies (Thompson *et al.*, 1991). In the case of cold agglutinins, this gene segment is responsible for the cross-reactive idiotype specificity characterisitic of the I/i response and is the primary gene segment found to encode this antigen reactivity in patients with cold agglutination disease (Pascual *et al.*, 1991, 1992*b*; Silberstein *et al.*, 1991; Grillot-Courvalin *et al.*, 1992; Leoni *et al.*, 1991). One might argue that the restriction is a direct result of the homogeneous nature of the antigen, i.e. carbohydrate. However, in a group of *structurally distinct* blood group antigens, including the A antigen, the Rh C, c, D, E, e and G antigens, and the Kidd antigens Jka and Jkb, this gene segment was found to represent 64% of the IgM antibodies and 21% of the IgG antibodies (Thompson *et al.*, 1991). This suggested a possible restriction in the human anti-red blood cell alloantibody response as well.

In order to further address this issue, we have characterized the heavy and light chain variable region gene segments expressed by six human monoclonal alloantibodies which bind various human HLA class I and class II alloantigens (Pistillo *et al.*, 1986, 1987, 1988, 1989, 1991, 1993; Mazzoleni *et al.*, 1989, 1991). These EBV-transformed cell lines were derived from healthy volunteers that were repeatedly immunized with blood from other individuals who were HLA-disparate. We did not find the V4-34 gene segment expressed within this group of antibodies, instead, the results reflect what has been observed in the majority of circumstances, a predominance of the V_H3 heavy chain gene family and, to a lesser extent, the V_κ3 and V_λ2 light chain gene families.

MATERIALS AND METHODS

Isolation and characterization of anti-class I and -class II producing cell lines

The isolation and characterization of the human class I- and class II-reactive monoclonal cell lines has been described previously (Pistillo *et al.*, 1987, 1988, 1989, 1991, 1995; Mazzoleni *et al.*, 1989, 1991). Briefly, healthy volunteers were immunized by transfusing with aliquots of whole blood from HLA-disparate donors at weekly intervals. Mononuclear cells were then isolated from these volunteers and transformed with EBV. Culture supernatants were screened for reactivity against donor-derived B lymphocytes. The specificity of each cell line is listed in Table 1. Cell lines MP1, MP9 and MP12 produce monoclonal IgM/λ antibodies, while cell lines MP6, MP10 and MP14 produce monoclonal IgM/κ antibodies. Each cell line was derived from a different donor as follows: MP1, Z. L., MP6, P. G., MP9, O. O., MP10, F. G., MP12, M. A., MP14, V. F. The period of time between the last immunization and the sampling of the cells was as follows: MP1, 1 year; MP6, 4 years; MP9, 6 years; MP10, 5 years; MP12, 1 year; and MP14, 2 years.

Oligonucleotides

The oligonucleotide primers used in the PCR amplifications were synthesized on an Applied Biosystems DNA synthesizer (Foster City, CA). The oligonucleotide sequences are shown in Table 2. Degenerate positions are denoted by (X/X). Rearranged V_H gene segments were amplified using the family-specific leader sequence primers in conjunction with the HUMCμ constant region primer. Rearranged V_λ gene segments were amplified with degenerate framework 1 primers (Songsivilai *et al.*, 1990) in conjunction with the HUMCλ constant region primer, designed to amplify all of the human λ constant region gene segments. Rearranged V_κ gene segments were ampli-

Table 1. Summary of V_H and V_L gene usage in MP cell lines

Cell line	Specificity line	V_H gene family	Closest germline	Per cent homology	J_H gene segment	V_L gene family	Closest germline	Per cent homology	J_L gene segment
MP1	DQB1*0201	V_H3	V3–13	95.5%	J_H2	$V_\lambda3$	VIII.1	95.7%	$J_\lambda3$
MP6	A*3002	V_H3	V3–23	97.6%	J_H3	$V_\kappa1$	VK02/012	99.7%	$J_\kappa2$
MP9	A*03	V_H1	V1–18	97.6%	J_H3	$V_\lambda2$	DPL10	97.2%	$J_\lambda3$
MP10	DRB1*01 DRB1*09 DRB1*1001 DRB1*15 DRB1*16	V_H4	VH4.18	96.6%	J_H4	$V_\kappa3$	L16	96.9%	$J_\kappa4$
MP12	DRB1*11	V_H3	M72	96.9%	J_H5	$V_\lambda2$	hslv2046	98.0%	$J_\lambda3$
MP14	DRB1*08 DRB1*12	V_H3	56p1	96.6%	J_H5	$V_\kappa3$	A27	97.6%	$J_\kappa2$

Table 2. Oligonucleotide sequences

Name	Sequence 5′→3′	Specificity
LVH1	ATGGACTGGACCTGGAGGATC	Primes V_H1 gene segments from the 5′ end of the leader
LVH3	CTCACCATGGAGTTTGGGCTG	Primes V_H3 gene segments from the 5′ end of the leader
LVH4	ATGAAACACCTGTGGTTC	Primes V_H4 gene segments from the 5′ end of the leader
HUMCμ	AAGGGTTGGGGCGGATGCACTCCC	Specific for the 5′ region of human MU constant region
LVK1	ATGGACA(T/C)GA(G/T)GG(T/C)C(C/T)C(G/A)CTCAG ATGGACAC(A/C)AGAG(T/C)CC(T/C)C(G/A)	A set of 2 degenerate oligonucleotides which prime $V_\kappa1$ GNE segments from the 5′ end of the leader
LVK3	ATGGAAACCCCAGCGCAG ATGGGGTCCCAGGTTCAC	A set of 2 oligonucleotides which prime $V_\kappa3$ gene segments from the 5′ end of the leader
HUMCκ	GACAGATGGTGCAGCCAC	Specific for the 5′ region of human κ constant region
VLAM1	CA(C/G)TCTCAGCTGAC(G/T)CA(A/G)CC(T/C/A/G)(C/G)CCTC	Set of three degenerate oligonucleotides which prime V_1 gene segments from the beginning of framework 1
VLAM2	TC(C/G)TATCAGCTGAC(G/T)CA(A/G)CC(T/C/A/G)CCCTC	
VLAM3	AATTTTCAGCTGAC(G/T)CA(A/G)CC(T/C/A/G)CACTC1	
HUMCλ	TGTGGCCTTGTTGGCTTGAAG	Specific for all human λ constant region gene segments

fied with the family-specific leader sequence primers in conjunction with the HUMCk constant region primer. Due to the large size of the $V_\kappa1$ and $V_\kappa3$ gene families, the leader sequences fall, roughly, into two groups, therefore, two primers were designed for amplification of members of these families.

First-strand cDNA synthesis

Total RNA was extracted from a frozen cell pellet by using RNA STAT-60 (Tel–Test "B" Inc., Friendswood, TX). This is a single-step isolation based on the methods of Chomczynski and Sacchi (1987) and Kedrierski and Porter (1991) which utilizes phenol and guanidinium thiocyanate in a monophase solution. First-strand cDNA synthesis was performed using oligo-dT as the primer and high concentration AMV reverse transcriptase (Promega Corp., Madison, WI) according to a modified protocol of Gubler and Hoffman (Ausubel *et al.*, 1989).

Polymerase chain reaction

Polymerase chain reactions (PCR) were performed essentially via the method recommended by Perkin–Elmer Cetus. Two of 30 μl of cDNA (nanogram quantities) were added to a 200 mM solution of each dATP, dCTP, dGTP and dTTP, with 600 ng of each primer and 2.25 U of *Taq* DNA polymerase (Promega Corp., Madison, WI). PCR cycles consisted of the following conditions: one cycle of denaturation at 94°C for 4 min; annealing at 55°C for 2 min; extension at 72°C for 2 min; 39 cycles of denaturation at 94°C for 1 min; annealing at 55°C for 2 min; and extension at 72°C for 2 min. Ampli-

fications were carried out in a Programmable Thermal Controller (MJ Research Inc., Watertown, MA).

Isolation, cloning and sequencing of the amplified products

Amplified products were size selected on a 1% agarose gel, ligated into the *EcoRV* site of Bluescript phagemid vector (Stratagene, La Jolla, CA), transformed into $CaCl_2$ competent XL-1 Blue bacteria (Stratagene) and screened by the blue/white colony screening method. Miniprep DNA was prepared from white colonies utilizing the Wizard Miniprep system (Promega Corp., Madison, WI) and digested with *Xba I* and *Xho I* to determine the size of the ligated product. ssDNA was generated from positive clones following superinfection with M13K07 helper phage (Biorad, Hercules, CA) as described previously (Pascual *et al.*, 1990). Sequencing was carried out on a minimum of four to six clones, from a single amplification, using the dideoxy chain termination method (Sanger *et al.*, 1977) with a modified version of T7 DNA polymerase (Sequenase, United States Biochemical, Cleveland, OH) (68) and the M13 universal primer. Gel electrophoresis was performed using Long Ranger gel solution (J.T. Baker Inc., Phillipsburg, NJ). Plasmid DNA containing the heavy and light chain variable region genes from cell lines MP6, MP12 and MP14 was also sequenced on an Applied Biosystems automated sequencer, model 373A (Applied Biosystems, Foster City, CA) using both the M13 universal primer and the M13 reverse primer. Completed sequences were analysed using DNASTAR (DNASTAR Inc., Madison, WI) and the combined EMBL/Genbank database.

RESULTS

Properties of anti-HLA cell lines and monoclonal antibodies

Six monoclonal B lymphoblastoid cell lines have been generated after EBV-immortalization of *in vivo* HLA-sensitized human B lymphocytes followed by repeated selection and subcloning of anti-HLA specific antibody-secreting cells. All of the cell lines that were obtained secrete allo-antibodies whose HLA specificity was analysed by microlymphocytotoxicity and cytofluorimetry against a panel of well-characterized HLA-homozygous B lymphoblastoid cell lines, HLA-typed peripheral blood B lymphocytes and HLA-transfected murine L cells. Segregation analysis within informative families has also been carried out.

In contrast to murine monoclonal antibodies, all of these human monoclonal antibodies recognize polymorphic HLA specificities, including class I and class II specificities (Table 1). Therefore, they are useful tissue typing reagents, in particular MP1, the first published monoclonal antibody identifying the celiac disease-associated DQB1*0201 allele; MP10, which allows identification of the rare DRB1*1001 phenotype; and MP6, which identifies the HLA-A*3002 (Pistillo *et al.*, 1995).

The availability of multiple HLA class II β chain first domain nucleotide and amino acid sequences allowed us to correlate the reactivity pattern of the anti-class II monoclonal antibodies with the presence of specific amino acid residues or clusters of residues that can be involved in the polymorphic epitopes recognized by the anti-class II monoclonal antibodies. This correlation demonstrated that amino acid residues unique to the DQB1 or DRB1 chains can be most likely involved in the formation of the epitopes recognized by the serologically monospecific antibodies MP1 and MP12. For the antibodies recognizing more than one HLA specificity, such as MP10 and MP14, amino acid residues shared by different DRB1 chains can contribute to the formation of the antibody binding site. In the case of MP12 the critical role played by glutamic acid at position 58 in the DRB1 chain has been clearly demonstrated by testing the ability to bind to wild-type or site-specific mutagenized transfectants (Klohe *et al.*, 1992).

The immunoglobulin isotype was determined by cytoplasmic immunofluorescence staining of the clones with immunoglobulin chain-specific antibodies. All of the clones produce IgM (up to 1 μg/ml of culture supernatant), but utilize different light chains and express members of different V_H and V_L families (see below).

Heavy chain variable regions

The complete nucleotide sequence of each heavy chain variable region is shown in Fig. 1. The V_H gene segment expressed by MP1 is a member of the V_H3 gene family. It is 95.5% identical to the germline V_H3 gene segment, V3-13, which is equivalent to DP48, V_H13–2 and 38p1 (Table 1: Matsuda *et al.*, 1993; Chothia *et al.*, 1992; Berman *et al.*, 1988; Schroeder *et al.*, 1987). The second cell line, MP6, also utilizes a V_H3 gene segment which demonstrates the closest sequence homology to the V3–23 (equivalent to DP47 and V_H26) germline gene segment (97.6%, Table 1; Matsuda *et al.*, 1993; Chothia *et al.*, 1992). The V_H gene segment utilized by MP9 is a member of the V_H1 gene family and displays 97.6% sequence homology to the germline gene segments, V1–18 (DP14; Table 1: Matsuda *et al.*, 1993; Chothia *et al.*, 1992). MP10 expresses a V_H4 gene segment that is most closely related to the germline gene segment, $V_H4.18$ (Table 1; Sanz *et al.*, 1989). The remaining two cell lines express members of the V_H3 gene family, distinct from each other and from those utilized by cell lines MP1 and MP6. MP12 expresses a V_H3 gene segment which is most closely related to another rearranged gene segment, M72, found in a fetal liver cDNA library (96.9%, Table 1; Schroeder and Wang, 1990). MP14 is 96.6% identical to another V_H3 gene segment commonly found in the fetal repertoire, 56p1 (Table 1; Schroeder *et al.*, 1987).

The comparison between each expressed V_H gene and the closest germline (or rearranged) gene segment is shown in Fig. 2. When each of these expressed sequences is compared to the most closely related germline or rearranged gene segment, the percent nucleotide difference ranges from 2.4 to 4.5%. These nucleotide differences are mainly distributed between complementarity

(a) MP1 VH

```
<----------LEADER----------> <----------
ATG GAG TTT GGG CTG AGC TGG GTT TTC CTT GTT GCT ATA TTA GAA GGT GTC CAG TGT GAG GTG CAG CTG GTG GAG
 M   E   F   G   L   S   W   V   F   L   V   A   I   L   E   G   V   Q   C   E   V   Q   L   V   E

----------VH3----------***
TCT GGG GGA GGC TTG GTA CAG CCT GGG GGG TCC CTG AGA CTC TCC TGT GCA TCC TCT GGA TTC ACC TTC AAT ACT
 S   G   G   G   L   V   Q   P   G   G   S   L   R   L   S   C   A   S   S   G   F   T   F   N   T

***CDR1********----------**********
TAC GAC ATG CAC TGG GTC CGC CAA GCT ACA GGA AAG GGT CTG GAG TGG GTC TCA GGT ATT CGT ACT GCT GGT GAC
 Y   D   M   H   W   V   R   Q   A   T   G   K   G   L   E   W   V   S   G   I   R   T   A   G   D

*********CDR2**********----------
ACA TAC TAT CCA GGC TCC GTG AAG GGC CGA TTC ACC ATC TCC AGA GAA AAT GCC AAG AAC TCC TTA TAT CTT GAA
 T   Y   Y   P   G   S   V   K   G   R   F   T   I   S   R   E   N   A   K   N   S   L   Y   L   E

----------> <----------D----------
ATG AAC ACT TTG AGA GCC GGG GAC AGG GCT GTG TAT TAC TGT GCA AGA GAG ATG TTC GAT AGT AGA GGT CAT TAC
 M   N   T   L   R   A   G   D   R   A   V   Y   Y   C   A   R   E   M   F   D   S   R   G   H   Y

------> <----------JH2----------
GGT CCC TTC GAT CTC TGG GGC CGG TGC ACC CTG GTC ACT GTC TCG TCA
 G   P   F   D   L   W   G   R   C   T   L   V   T   V   S   S
```

(b) MP6

```
<----------LEADER----------> <----------
ATG GAG TTT GGG CTG AGC TGG CTT TTT CTT GTG GCT ATA TTA AAA GGT GTC CAG TGT GAG GTG CAG CTG TTG GAG
 M   E   F   G   L   S   W   L   F   L   V   A   I   L   K   G   V   Q   C   E   V   Q   L   L   E

----------VH3----------***
ACT GGG GGA GGC TTG GTG CAG CCT GGG GGG TCC CTG AGA CTC TCC TGT GCA GCC TCT GGA TTC ACC TTT AGC AGC
 T   G   G   G   L   V   Q   P   G   G   S   L   R   L   S   C   A   A   S   G   F   T   F   S   S

*****CDR1******----------**********
TAT GCC ATG AGC TGG GTC CGC CAG GCT CCA GGG AAG GGG CTG GAG TGG GTC TCA GCT ATT AGT GGT AGT GGT GGT
 Y   A   M   S   W   V   R   Q   A   P   G   K   G   L   E   W   V   S   A   I   S   G   S   G   G

*******CDR2**********----------
CAG ACA TAC TAC GCA GAC TCC GCG AAG GGC CGG TTC ACC ATC TCC AGA GAC AAT TCC AAG AAC ACG TTG TAT CTG
 Q   T   Y   Y   A   D   S   A   K   G   R   F   T   O   S   R   D   N   S   K   N   T   L   Y   L

----------> <----------D----------
CAA ATG AAC AGC CTG AGA GCC GAG GAC ACG GCC GTA TAT TAC TGT GCG AAA GAG AGG GGT TAC TAT GAT AGT CCG
 Q   M   N   S   L   R   A   E   D   T   A   V   Y   Y   C   A   K   E   R   G   Y   Y   D   S   P

--> <----------JH3----------
TAT GCT TTA GAT ATC TGG GGC CAA GGG ACA ATG GTC ACC GTC TCT TCA
 Y   A   L   D   I   W   G   Q   G   T   M   V   T   V   S   S
```

(c) MP9

```
<----------LEADER----------> <----------
ATG GAC TGG ACC TGG AGG ATC CTT TTC TTG GTG GCA GCA GCA ACA GGT GCC CAC TCC CAG GTT CAG CTG GTG CAG
 M   D   W   T   W   R   I   L   F   L   V   A   A   A   T   G   A   H   S   Q   V   Q   L   V   Q

----------VH1----------***
TCT GGA GCT GAG GTG AAG AAG CCT GGG GCC TCA GTG AAG GTC TCC TGC AAG GCT TCT GGT TAC ACC TTT ACC AGC
 S   G   A   E   V   K   K   P   G   A   S   V   K   V   S   C   K   A   S   G   Y   T   F   T   S

*****CDR1******----------**********
TAT GGC ATC AGC TGG GTG CGA CAG GCC CCT GGA CAA GGG CTT GAG TGG ATG GGA TGG ATC AGC GCC TAC AAT GGA
 Y   G   I   S   W   V   R   Q   A   P   G   Q   G   L   E   W   M   G   W   I   S   A   Y   N   G

***************CDR2**********----------
AAC ACA AAC TAT GCA CAG AAG CTC CAG GGC AGA GTC ACC ATG ACC ACA GAC ACA TCC ACG AGC ACA GCC TAC ATG
 N   T   N   Y   A   Q   K   L   Q   G   R   V   T   M   T   T   D   T   S   T   S   T   A   Y   M

----------> <----------D----------
GAG CTG AGG AGC CTG AAA TCT GAC GAC ACG GCC GTG TAT TAT TGT GGT AGA CAA TGG TTC GGG GAG TCG ATC TAC
 E   L   R   S   L   K   S   D   D   T   A   V   Y   Y   C   G   R   Q   W   F   G   E   S   I   Y

----------> <----------JH3----------
TAC TAC TAC ATG GAC GTC TGG GGC AAA GGG ACC ACG GTC ACC GTC TCC TCA
 Y   Y   Y   M   D   V   W   G   K   G   T   T   V   T   V   S   S
```

Fig. 1. cDNA sequence of the anti-HLA heavy chain variable regions. The leader, V_H, D and J_H gene segments are denoted by the dashed line; CDR1 and CDR2 are indicated by asterisks. (a) cDNA sequence of MP1; (b) cDNA sequence of MP6; (c) cDNA sequence of MP9; (d) cDNA sequence of MP10; (e) cDNA sequence of MP12; and (f) cDNA sequence of MP14. The sequences have been assigned Genbank accession numbers L38431, L38433, L38435, L38425, L38427 and L38429, respectively.

(d) MP10

```
<----------------------------LEADER-----------------------------------> <--------------------
ATG AAA CAC CTG TGG TTC TTC CTC CTG CTG GTG GCG GCT CCC AGA TGG GTC CTG TCC CAG CTG CAG CTG CAG GAG
 M   K   H   L   W   F   F   L   L   L   V   A   A   P   R   W   V   L   S   Q   L   Q   L   Q   E

--------------------VH4-------------------------------------------------------------------------***
TCG GGC CCA GGA CTC GTG AAG CCT TCG GAG ACC CTG TCC CTC ACC TGC ACT GTC TCT GTG GGC TCC ATT AGT AGT
 S   G   P   G   L   V   K   P   S   E   T   L   S   L   T   C   T   V   S   G   G   S   I   S   S

******CDR1**********----------------------------------------------------------+******************+
AGT AGT CAC TAC TGG GGC TGG ATC CGC CAG CCC CCA GGG AAG GGA CTG GAG TGG ATT GGG ACT ATC TAT TAT AGT
 S   S   H   Y   W   G   W   I   R   Q   P   P   G   K   G   L   E   W   I   G   T   I   Y   Y   S

**************CDR2**********************+-------------------------------------------------------
GGG AGC ACC TAC CAC AAC CCG TCC CTC AAG AGC CGA GTC ACC ATA TCC GTA GAC ACG TCC AAG AAC CAG TTC TCC
 G   S   T   Y   H   N   P   S   L   K   S   R   V   T   I   S   V   D   T   S   K   N   Q   F   S

-------------------------------------------------------------> <-----------D------------>
CTG AAG CTG AGC TCT GTG ACC GCC ACA GAC ACG GCT GTG TAT TTC TGT GCG AGA CAC CTC GGG CCC TGG GAA AAC
 L   K   L   S   S   V   T   A   T   D   T   A   V   Y   F   C   A   R   H   L   G   P   W   E   N

<------------------JH4-------------------->
TGG GGC CAG GGA ACC CTG GTC ACC GTC TCC TCA
 W   G   Q   G   T   L   V   T   V   S   S
```

(e) MP12

```
<--------------------------LEADER-----------------------------------> <--------------------
ATG GAG TTT GGG CTG AGC TGG GTT GTC CTC GTT GCT CTT TTA AGA GGT GTC CAG TGT CAG GTG CAG CTG GTG GAG
 M   E   F   G   L   S   W   V   V   L   V   A   L   L   R   G   V   Q   C   Q   V   Q   L   V   E

-------------------------VH3-------------------------------------------------------------------***
TCT GGG GGA GGC GTG GTC CAG CCT GGG AGG TCC CTG AGA CTC TCC TGT GCA GCC TCT GGA TTC TCC TTC AGC AGA
 S   G   G   G   V   V   Q   P   G   R   S   L   R   L   S   C   A   A   S   G   F   S   F   S   R

****CDR1*******----------------------------------------------------------+**************************
TAT GCT ATG TAC TGG GTC CGC CAG GCT CCA GGC AAG GGG CTG GAG TGG GTG GCA GTT ATA TCA TAT GAT GGA AGT
 Y   A   M   Y   W   V   R   Q   A   P   G   K   G   L   E   W   V   A   V   I   S   Y   D   G   S

**************CDR2**********************+-------------------------------------------------------
AAT AAA TAT TAT GCA GAC TCC GTG AAG GGC CGA TTC ACC ATC TCC AGA GAC AAT TCC AAG AAC ACG CTG TAT CTG
 N   K   Y   Y   A   D   S   V   K   G   R   F   T   I   S   R   D   N   S   K   N   T   L   Y   L

-------------------------------------------------------------> <----------------D------------
CAA ATG GAC AGC CTG AGA GCT GAC GAC ACG GCT GTG TAT TAC TGT GCG GGA GGA GTG GTT ATT ATA TTT AGT CGA
 Q   M   D   S   L   R   A   D   D   T   A   V   Y   Y   C   A   G   G   V   V   I   I   F   S   R

----------> <-------------------JH5--------------------->
CTT GAT TAC TGG GGC CAG GGA AAC CTG GCC ACC GTC TCC TCA
 L   D   Y   W   G   Q   G   N   L   A   T   V   S   S
```

(f) MP14

```
<----------------------------LEADER-----------------------------------> <--------------------
ATG GAG TTT GGG CTG AGC TGG GTT TTC CTC GTT GCT CTT TTA AGA GGT GTC CAG TGT CAG GTG CAA CTG GTG GAG
 M   E   F   G   L   S   W   V   F   L   V   A   L   L   R   G   V   Q   C   Q   V   Q   L   V   E

-------------------------------VH3-------------------------------------------------------------***
TCT GGG GGA GGC GTG GTC CAG CCT GGG AGG TCC CTG AGA CTC TCC TGT GCA GCC TCT GGA TTC ACC TTC ACT AGC
 S   G   G   G   V   V   Q   P   G   R   S   L   R   L   S   C   A   A   S   G   F   T   F   T   S

*****CDR1******----------------------------------------------------------+**************************
TAT GCT ATG CAC TGG GTC CGC CAG GCT CCA GGC AAG GGA CTG GAG TGG GTG GCA GTT ATG TCA TTT GAT GGA AGC
 Y   A   M   H   W   V   R   Q   A   P   G   K   G   L   E   W   V   A   V   M   S   F   D   G   S

**************CDR2**********************+-------------------------------------------------------
AAA AAA TAC TAC GCA GAC TCC GTG AAG GGC CGA TTC ACC ATC TCC AGA GAC AAT TCC AAG AAC ACA CTG TTT CTG
 K   K   Y   Y   A   D   S   V   K   G   R   F   T   I   S   R   D   N   S   K   N   T   L   F   L

-------------------------------------------------------------> <----D-----> <------------------
CAA ATG AAC AGC CTG AGA GCT GAG GAC ACG GCT ATT TAT TAC TGT GCG AGA GAT CAA ATG GGT TGG TTC GAC CCC
 Q   M   N   S   L   R   A   E   D   T   A   I   Y   Y   C   A   R   D   Q   M   G   W   F   D   P

----------------JH5------------------------>
TGG GGC CAG GGA ACC CTG GTC ACC GTC TCC TCA
 W   G   Q   G   T   L   V   T   V   S   S
```

Fig. 1– *continued.*

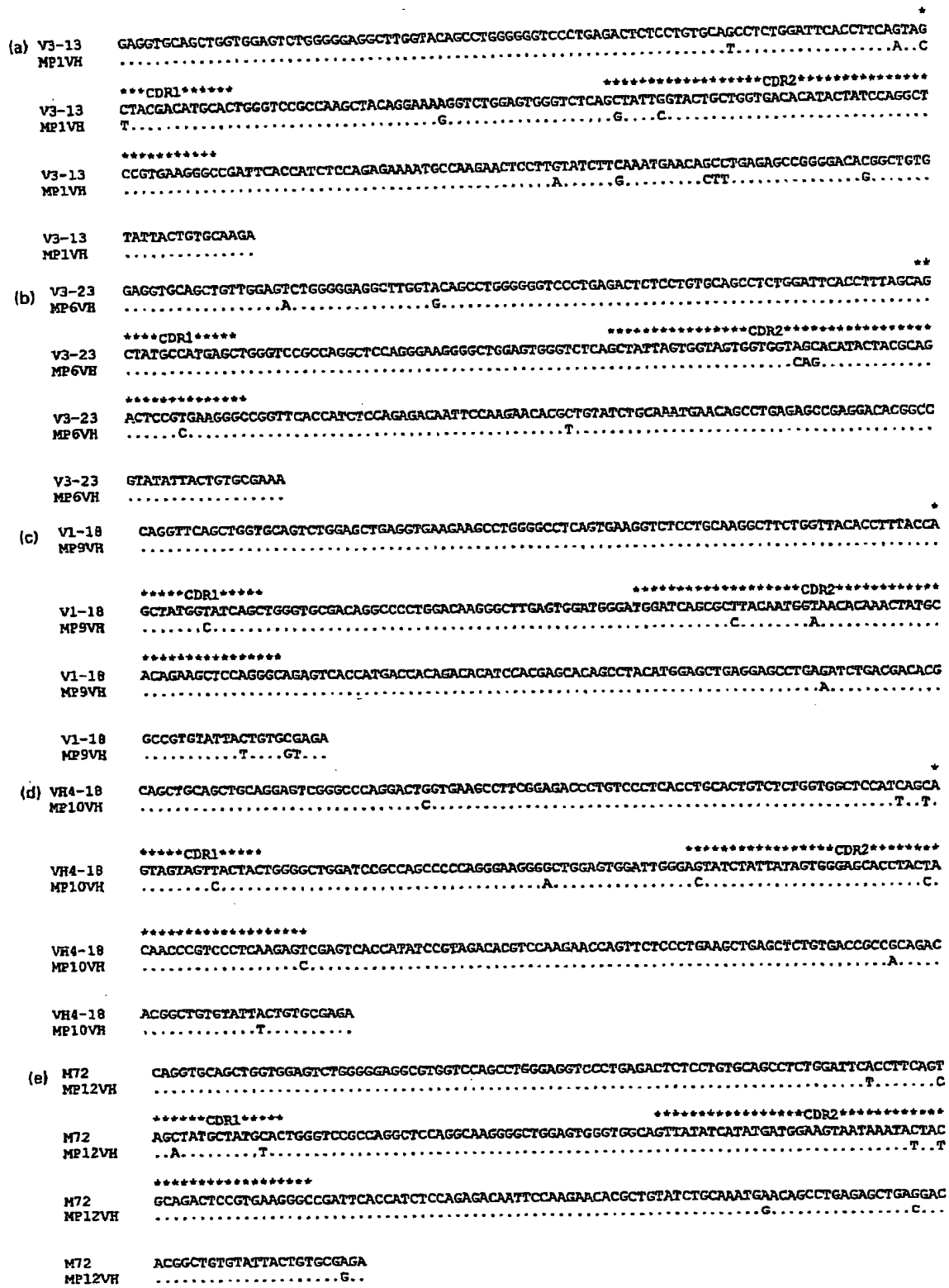


Fig. 2. Comparison of the nucleotide sequences of anti-HLA heavy chain variable regions and the closest germline or rearranged gene segment. Identity between the sequences is denoted by the periods; CDR1 and CDR2 are indicated by asterisks. (a) MP1VH vs V3–13 (Matsuda *et al.*, 1993); (b) MP6VH vs V3–23 (Matsuda *et al.*, 1993); (c) MP9VH vs V1–18 (Matsuda *et al.*, 1993); (d) MP10VH vs VH4.18 (Sanz *et al.*, 1989); (e) MP12VH vs M72 (Schroeder and Wang, 1990); and (f) MP14VH vs 56p1 (Schroeder *et al.*, 1987).

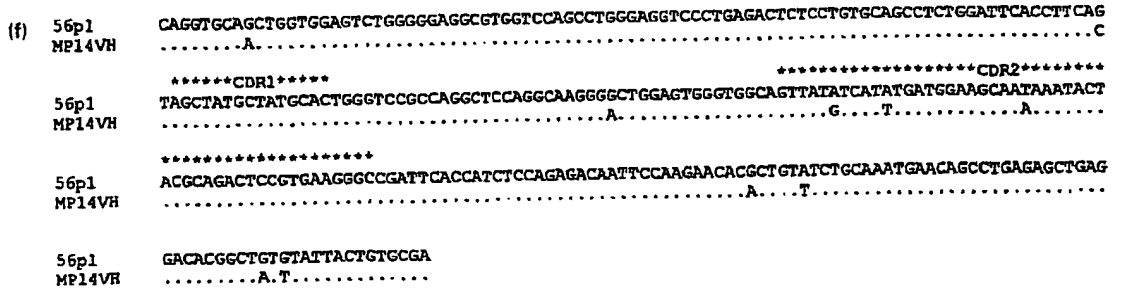


Fig. 2 – *continued.*

determining region 1 (CDR1), CDR2 and framework 3 (FR3), with occasional changes in FR1 or FR2. With the exception of MP9, in which the majority of nucleotide differences do not result in a change in amino acid (i.e. silent mutations), the remainder of the expressed V_H gene segments exhibit a ratio of replacement to silent substitutions ranging from 2:1 to 7:1 (Table 3). Furthermore, the majority of the replacement substitutions fall within CDRs and FR3, suggestive of an antigen-driven selection of these B cells. Overall, silent nucleotide changes are rare, but fall mainly within framework regions, which may be more reflective of polymorphism than somatic mutation. These observations are not unique in the study of human immunoglobulin variable region gene segments, as shown by the numerous aforementioned studies. The main difference between the current observations and our previous observations is the overall extent of apparent somatic mutation. In a group of human IgG antibodies that are specific for a number of different exogenous antigens, the percent nucleotide difference with a corresponding germline gene segment is generally more extensive than observed with these IgM antibodies. This observation may reflect the results of a recent study in which various tonsillar B cell subsets were assessed for the amount of somatic mutation that had been introduced into the expressed variable region gene segments (Pascual *et al.*, 1994). It was clearly shown that there are naive subsets of B cells, which express only surface IgM and IgD, and express V_H gene segments that are essentially germline in origin; while there are other IgM-expressing subsets (germinal center) that have begun to accrue point mutations. These appear to preceed the IgG-expressing, germinal center subset that have accumulated a substantial number of somatic mutations. On average, the number of base-pair substitutions among the IgG transcripts from germinal center-derived cells was two-fold compared to the IgM transcripts from the same B cell subpopulation.

Light chain variable regions

The complete light chain variable region sequence of each cell line is shown in Fig. 3. Three of the cell lines express light chains of the λ isotype, while four express κ light chains (Table 1). The variable region gene expressed in the MP1 cell line is a member of the $V_\lambda 3$ gene family. It exhibits the closest homology to the germline gene, VIII.1, which is equivalent to DPL23 (95.7%, Table 1; Combriato and Klobeck, 1991; Williams and Winter, 1993). The second cell line, MP6, utilizes a V_κ gene segment belonging to the $V_\kappa 1$ family. This expressed sequence is 99.7% identical to the $V_\kappa 02/012$ gene segments, two identical germline gene segments which have been duplicated within the human κ locus (Table 1; Pargent *et al.*, 1991). The third cell line, MP9, expresses a V_λ gene which is a member of the $V_\lambda 2$ gene family. This expressed sequence displays 97.2% sequence identity to the germline gene, DPL10 (equivalent to hslv2066; Table 1; Williams and Winter, 1993; Irigoyen *et al.*, 1994). MP10 appears to utilize the germline gene segment, L16 (Humkv328), with a sequence identity of 96.9% (Table 1; Huber *et al.*, 1993; Liu *et al.*, 1989). Like MP9, MP12 also expresses a member of the $V_\lambda 2$ family, distinct from that expressed in the MP9 cell line. MP12V_λ displays the closest sequence homology with the germline $V_\lambda 2$ gene, hslv2046 (98%, Table 1; Irigoyen *et al.*, 1994). The last

Table 3. Summary of nucleotide differences between expressed and germline V_H gene segments in anti-HLA antibodies

	FR1		CDR1		FR2		CDR2		FR3	
Cell line	R/S	% MUT	R/S	% MUT	R/S	%MUT	R/S	% MUT	R/S	% MUT
MP1	2/0	2.2%	2/0	13.3%	0/1	2.4%	2/0	4.2%	6/1	7.3%
MP6	1/1	2.2%	0/0	0.0%	0/0	0.0%	4/0	7.8%	1/0	1.0%
MP9	0/0	0.0%	0/1	6.7%	0/0	0.0%	0/2	3.9%	3/1	4.2%
MP10	3/0	3.3%	1/0	4.8%	0/1	2.4%	2/1	6.3%	2/0	2.1%
MP12	1/1	2.2%	2/0	13.3%	0/0	0.0%	0/2	3.9%	3/0	3.1%
MP14	1/1	2.2%	0/0	0.0%	0/1	2.4%	3/0	5.9%	3/1	4.2%

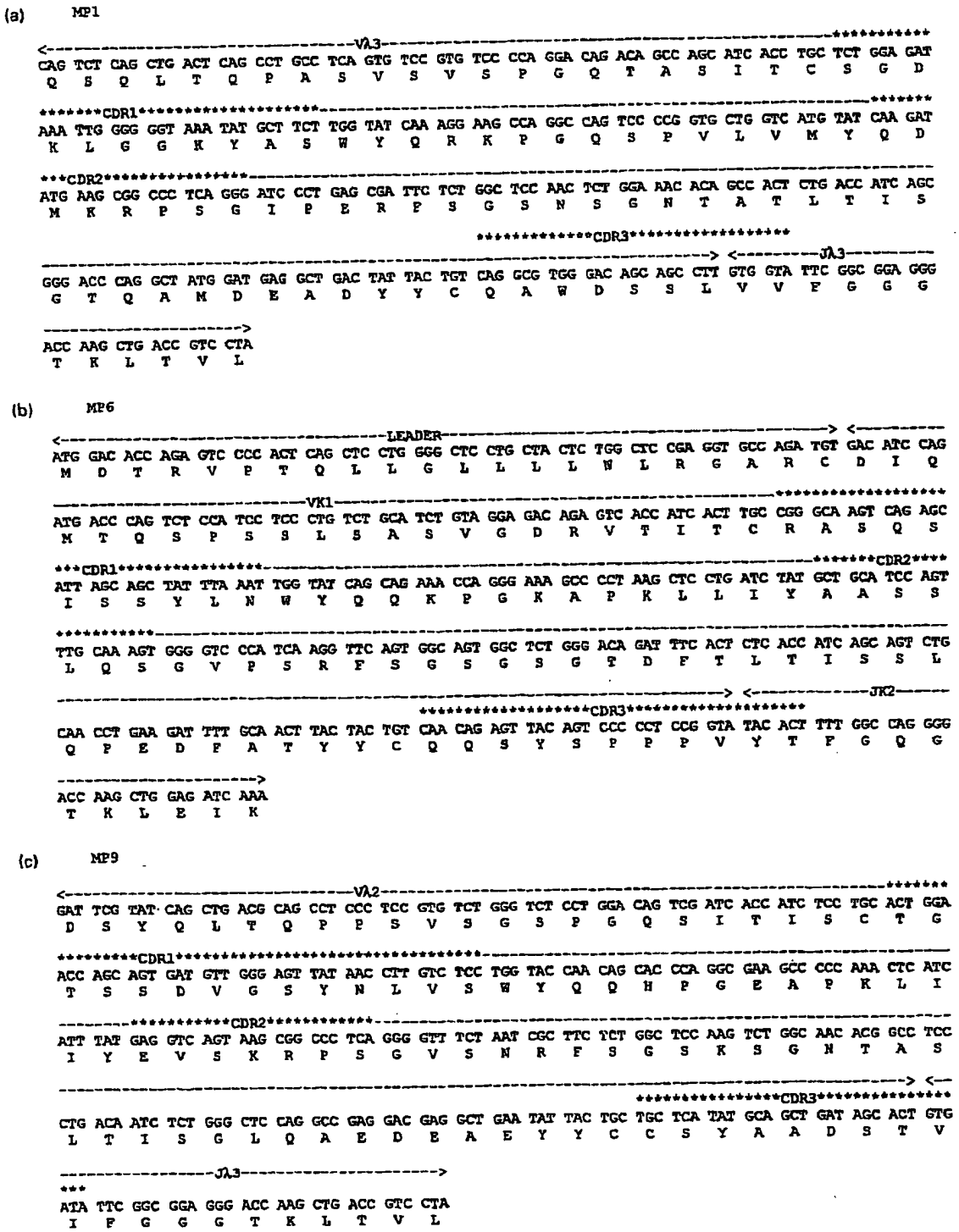
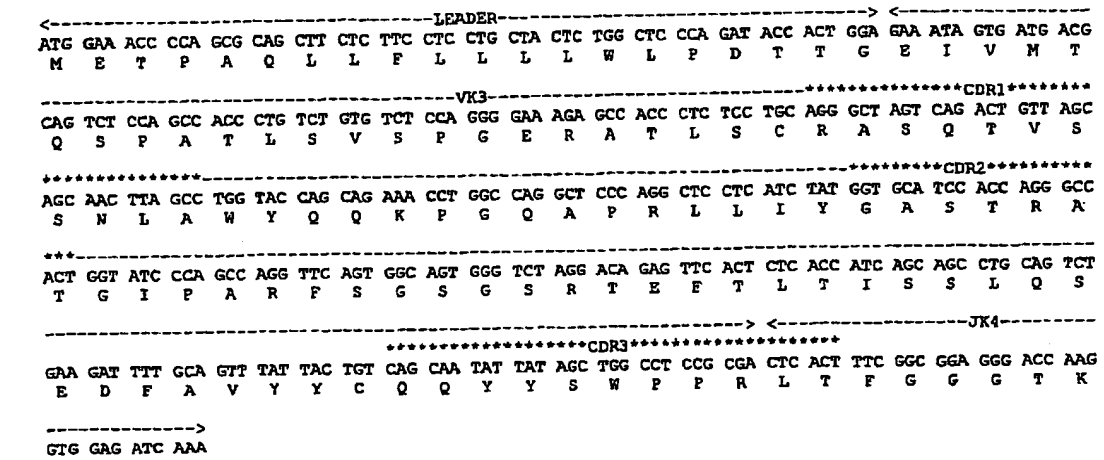
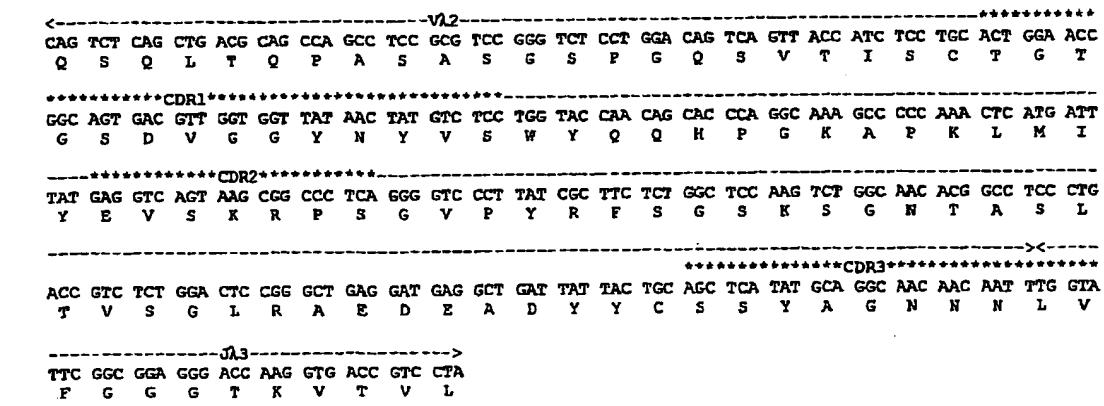


Fig. 3. cDNA sequences of anti-HLA light chain variable regions. The leader, V_H, D and J_H are denoted by the dashed line; CDR1, CDR2 and CDR3 are indicated by asterisks. (a) cDNA sequence of MP1; (b) cDNA sequence of MP6; (c) cDNA sequence of MP9; (d) cDNA sequence of MP10; (e) cDNA sequence of MP12; and (f) cDNA sequence of MP14. The sequences have been assigned Genbank accession numbers L38432, L38434, L38436, L38426, L38428 and L38430, respectively.

(d) MP10



(e) MP12



(f) MP14

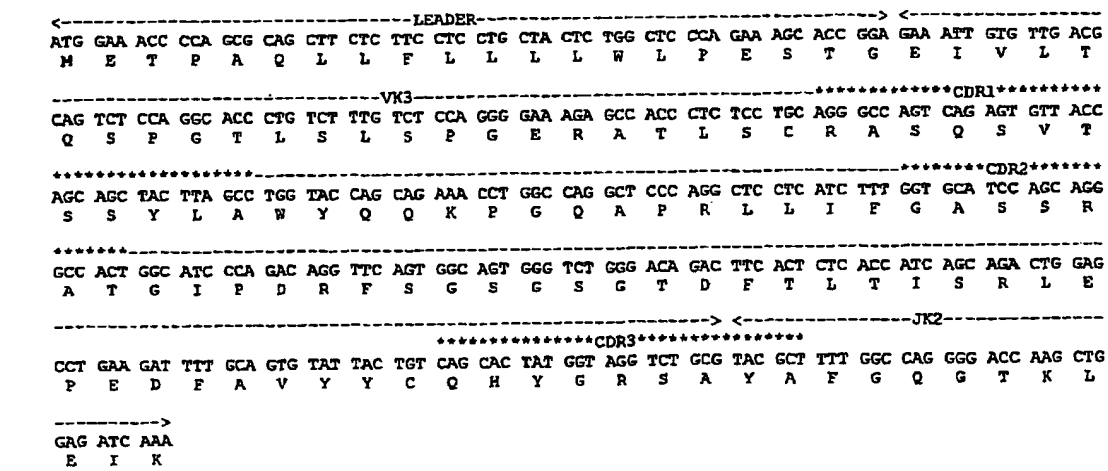


Fig. 3—continued.

cell line, MP14, expresses a $V_\kappa 3$ gene segment which is 97.6% identical to the germline gene, A27 (Table 1; Straubinger *et al.*, 1988).

The comparison between each of the expressed V_L genes and the closest germline gene is shown in Fig. 4. When each of the expressed V_λ gene segments is compared to a putative germline sequence, the percent nucleotide difference ranges from 1.1 to 4.2% (the differences in the beginning of framework 1 in the λ genes are likely to be due to the degeneracy of the PCR primers and have not been considered in the homology analysis). As seen in Table 4, the overall ratio of replacement to silent substitutions ranges from 2:1 to 8:1. Similar to the heavy chain, this is indicative of antigen-driven selection. In general CDR3 exhibits the majority of the changes, while CDR1 and CDR2 demonstrate fewer than that observed in the heavy chain.

Two of the three κ-expressing cell lines appear to have additional nucleotides at the V–J junction. It was originally thought that length variation in light chains was a result of recombination slippage rather than N segment addition, as terminal deoxynucleotidyl transferase (TdT) activity is negligible at the time of light chain rearrangement. However, it has been demonstrated that N segment addition can contribute to overall light chain diversity (Victor and Capra, 1994). In comparing $MP6V_\kappa$ to the germline $V_\kappa 02/012$ gene segment, four nucleotides at the end of the V gene do not appear to be derived from the germline sequence (Fig. 4B). It has been suggested that one type of junctional diversity, P nucleotides, can be attributed to the transfer of the terminal nucleotide(s) from the antisense strand to the sense strand during recombination, forming a palindromic template (Lafaille *et al.*, 1989). The first two nucleotides, GG, may be the result of P nucleotides, derived from the complementary CC at the 3′ end of the germline gene. However, the last two nucleotides, TA, cannot be attributed to either the V gene recombination signal sequence (RSS) or the J gene segment. Therefore it is possible that these may be the result of TdT activity. $MP10V_\kappa$ also exhibits four additional nucleotides at the V–J junction (Fig. 4D). The first G, again, may be the result of P nucleotide addition, derived from the complementary terminal C. However, the CGA cannot be attributed to either the V gene segment or the J gene segment or their corresponding RSS. Again, they may be the result of N segment addition by TdT. However, as with the heavy chain variable region, one must use caution in assessing the germline origin, somatic mutation, and junctional diversity, as one cannot be certain of polymorphism and undescribed germline gene segments.

D and J gene segments

In the heavy chain variable region the D segment comprises the majority of CDR3 and is believed to be responsible for much of the antibody specificity. With few exceptions, the assessment of D segment utilization in human immunoglobulins has been, and continues to be, difficult, due to the generally poor homology when comparing expressed D segments and the known germline D segments. Similarly, our D segment analysis has revealed limited homology between any of the expressed sequences and the known germline sequences. It has been demonstrated that D segments can be found expressed in the forward and reverse orientations, as well as fused to each other in both orientations (Meek *et al.*, 1989; Sanz, 1991; Tuaillon *et al.*, 1993). One example of a possible D–D fusion is found with the MP1 D segment. The expressed sequence is homologous to two fetal D gene segments, D21-7 (Sanz, 1991) at the 5′ end and D21-9 (Sanz, 1991) at the 3′ end, with flanking nucleotides at both ends which do not exhibit significant homology to any germline sequence (comparison not shown). The first 15 nucleotides of the MP12 D segment are identical to the germline DXP4 gene segment (Sanz, 1991), leaving 10 nucleotides without significant homology to any other known D gene (comparison not shown). The profile of this stretch of nucleotides is uncharacteristic of N segment addition, as it is A/T-rich, therefore it may represent an undescribed germline sequence. Although the expressed D segments range in size from nine to 42 nucleotides, no other significant homologies were found.

In addition, without a definite germline D segment donor and V gene flanking sequence, it becomes difficult to analyse the derivation of individual nucleotides. Many of the known germline V gene segments have been reported in the literature as PCR-generated fragments which do not include any flanking sequence that may become part of the expressed sequence during rearrangement. In the case of MP6, the first two nucleotides of the defined D segment (GA) may have derived from the two nucleotides that are present in the germline sequence between the end of the coding sequence and the RSS. There does not appear to be any P nucleotide addition at either the V–D or the D–J junction in any of the heavy chain variable regions.

Among these antibodies, one expresses J_H2, one expresses J_H4, two express J_H3 and two express J_H5. This differs from previous studies where J_H4 has been shown to predominate in the normal adult antibody repertoire, while J_H3 and J_H5 are significantly decreased (Table 1; Tuaillon *et al.*, 1993; Yamada *et al.*, 1991).

With the exception of MP6, all of the antibodies exhibit signs of exonuclease activity at the 5′ end of the J_H segment, i.e. the homology between the expressed J_H and the germline J_H begins approximately three to five nucleotides from the 5′ end of the germline sequence. All of the expressed J_H gene segments display nucleotide differences when compared to the corresponding germline sequences.

In the λ-expressing clones, all three utilize the $J_\lambda 3$ gene segment. This finding is similar to our previous studies in which $J_\lambda 2$ and $J_\lambda 3$ predominate among λ light chains. In contrast to the heavy chain J segments, the expressed J_λ genes do not appear to have undergone exonuclease digestion. In addition, MP1 J_λ is completely germline, while MP9 and MP12 have only one nucleotide difference each, compared to the germline $J_\lambda 3$. In the κ-expressing clones, one expresses $J_\kappa 4$ and two express $J_\kappa 2$. As with the λ light chains, three of the κ light chains exhibit no

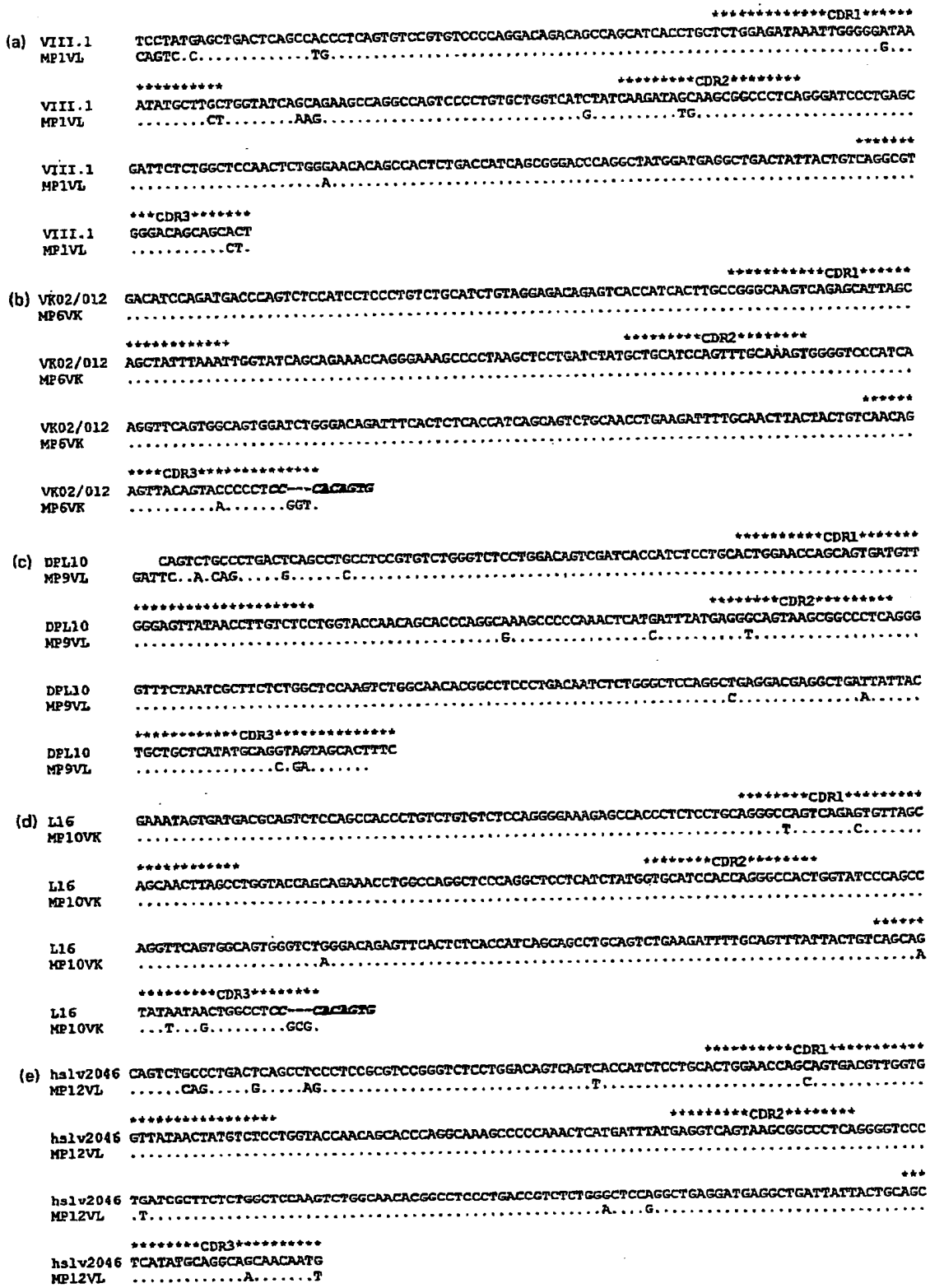


Fig. 4. Comparison of the nucleotide sequences of anti-HLA light chain variable regions with the closest germline gene segment. Identity between the sequences is indicated by the periods; CDR1, CDR2 and CDR3 are denoted by the asterisks; the RSS and additional, non-coding germline nucleotides are shown in italic. (a) MP1VL vs VIII.1 (Combriato and Klobeck, 1991); (b) MP6VL vs O2/O12 (Pargent *et al.*, 1991); (c) MP9VL vs DPL10 (Williams and Winter, 1993); (d) MP10VL vs L16 (Huber *et al.*, 1993); (e) MP12VL vs hslv2046 (Irigoyen *et al.*, 1994); and (f) MP14VL vs A27 (Straubinger *et al.*, 1988).

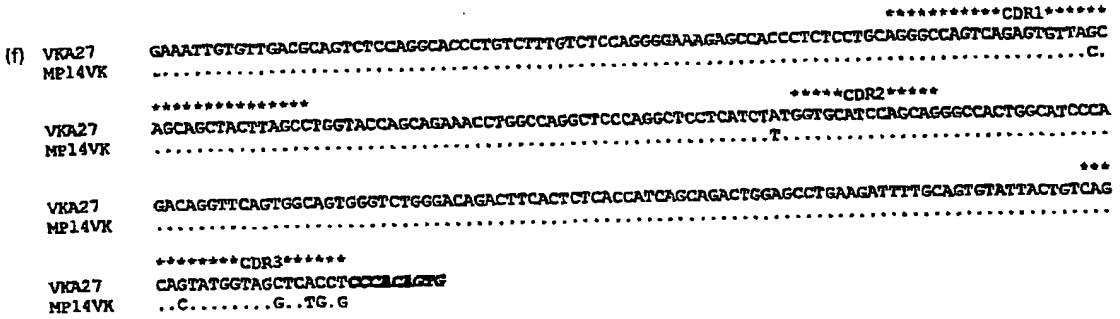


Fig. 4—*continued.*

signs of exonuclease activity at the 5′ end of the J segment. Two of these three J_κ segments are germline (MP6 and MP14) and the third has only one nucleotide difference compared to the germline sequence (MP10).

DISCUSSION

Currently, the data regarding the repertoire of variable region gene segments utilized in response to alloantigens is limited to a single study involving a substantial group of monoclonal antibodies specific for a number of structurally distinct, human blood group antigens (Thompson *et al.*, 1991). It was found that a significant proportion of these antibodies utilized a single gene segment, V4–34 (V_H4–21). To further examine the alloantibody repertoire, we have sequenced, at the nucleotide level, the heavy and light chain variable region gene segments expressed by six human IgM monoclonal alloantibodies which are specific for various HLA class I and class II molecules. From these results we conclude the following: (1) the distribution of V_H and V_L gene families reflects what has been described previously, i.e. a predominance of the V_H3 gene family, and to a lesser extent, the V_κ3 and V_λ2 gene families; (2) there is not a restriction in the expression of individual *gene segments*; and (3) in general, there is minimal apparent somatic mutation, compared to our previous studies describing antibodies to a variety of other exogenous antigens.

As mentioned previously, due primarily to the prevalence of V region CRI-association with different autoimmune diseases, it has been hypothesized that different gene segments are maintained in the germline for use in different types of responses. The interpretation of the data regarding the characterization of variable region gene segment usage in a number of different responses to both exogenous and endogenous antigens would suggest that the majority of germline variable region gene segments can be utilized for virtually any antigen specificity (for references, see introduction). One exception may be the V4–34 (V_H4–21) gene segment (Pascual and Capra, 1992). This gene segment has been shown to encode the CRI, 9G4, expressed on human cold agglutinins with the I/i specificity (Stevenson *et al.*, 1986). Although this gene segment is highly represented in fetal and adult lymphoid tissues, it is found encoding only 0.2 and 0.6% of normal serum IgM and IgG protein, respectively (Pascual and Capra, 1992). Thus far, V4–34 has only been found to be utilized in the autoantibody and anti-blood group alloantibody pools, presumably representing some fraction of the aforementioned serum IgM and IgG. It has only been found in one of 30–40 antibodies specific for other exogenous antigens (Bogard *et al.*, 1993). In a study of a large group of human alloantibodies reactive to a diverse array of structurally distinct blood group antigens, V4–34 was found to be expressed by 64% of IgM and 21% of IgG expressing clones (Thompson *et al.*, 1991). Since the panel of antigen specificities was diverse, both in number and structure, this suggested that V4–34 may be responsible for a significant portion of alloantibody responses. The current study indicates that a number of different gene segments are recruited in response to other alloantigens, consequently, the previously observed usage of V4–34 in anti-blood group antibodies may be the result of something inherent to the structure of the

Table 4. Summary of nucleotide differences between expressed and germline V_L gene segments in anti-HLA antibodies

	FR1		CDR1		FR2		CDR2		FR3		CDR3	
Cell line	R/S	% MUT	R/S	% MUT	R/S	%MUT	R/S	% MUT	R/S	% MUT	R/S	%MUT
MP1	0/0	0.0%	3/0	9.1%	3/1	8.9%	0/2	9.5%	0/1	1.0%	2/0	9.5%
MP6	0/0	0.0%	0/0	0.0%	0/0	0.0%	0/0	0.0%	0/0	0.0%	3/1	16.7%
MP9	1/0	1.4%	0/0	0.0%	2/0	4.4%	1/0	4.8%	1/1	2.1%	3/0	12.5%
MP10	0/0	0.0%	1/1	6.1%	0/0	0.0%	0/0	0.0%	1/0	1.0%	2/1	14.3%
MP12	0/1	1.4%	1/0	2.4%	0/0	0.0%	0/0	0.0%	2/1	3.1%	1/0	4.2%
MP14	0/0	0.0%	1/0	2.8%	1/0	2.2%	0/0	0.0%	0/0	0.0%	3/2	23.8%

red blood cell itself or the structure of its antigens that is currently unknown. The results of this study show that the distribution of variable region gene families and the ratio of κ and λ light chain-bearing antibodies generally reflect the findings of others, i.e. a predominance of the V_H3 heavy chain gene family and the $V_\kappa3$ and $V_\lambda2$ light chain gene families. However, due to the sample size and the diverse array of specificities, no correlations can be made between V genes and HLA polymorphisms.

Unlike our previous studies, none of the antibody V regions described here express gene segments derived from single (or two) member families, therefore it is difficult to precisely determine the extent of somatic mutation. Although there is now extensive germline sequence data (Matsuda *et al.*, 1993; Cook *et al.*, 1994), the true extent of polymorphism in individual gene segments is still unknown, thereby allowing only assessment of *apparent* somatic mutation by comparing expressed sequences with a given germline or rearranged gene segment that displays the most significant nucleotide homology. This analysis suggests that these expressed heavy and light chain variable regions are somewhat less mutated than many of the previously described sequences, some of which exhibited as much as 15% disparity from the closest germline gene segment (Andris *et al.*, 1991). This observation may be reflective of two recent studies of the extent of somatic mutation in different subsets of B lymphocytes representing various stages of maturation (Pascual *et al.*, 1994; Kuppers *et al.*, 1993). It was found that the IgM^+, IgD^+ naive B cells express variable region gene segments that are germline in origin, with no detectable γ transcripts, whereas germinal center (antigen-stimulated) B cells fall into two groups with regard to somatic mutation. The first group contains those cells which still express IgM and exhibit low to moderate somatic mutation, while the second group includes the IgG-expressing transcripts that have a more extensive number of mutations. The IgM alloantibodies described in this study may be representative of the IgM antibodies in the germinal center, while previously described IgG antibodies may represent the second group of germinal center cells that have undergone isotype switching and more extensive mutation and selection. More precisely, they may represent B lymphocytes from the memory cell pool, as the length of time from the last immunization to the time of sampling ranges from 1 to 6 years. This assessment still fits the scheme set forth by Pascual *et al.*, as they describe the IgM-expressing memory cell pool to exhibit an average of 3–6 nucleotide substitutions. Although IgG-bearing cells appear to predominate in the memory cell population, the IgM positive memory cells may represent those cells that have not undergone extensive somatic mutation and class switching, both of which may require a more prolonged exposure to antigen. Alternatively, these cells may have been sampled from the pool of newly generated cells, and therefore, represent IgM^+ cells that have been recently recruited, rather than recirculating memory cells.

Nonetheless, taken as a whole, the nucleotide differences observed in these anti-HLA antibodies are characteristic of cells which have undergone some amount of somatic hypermutation and antigenic selection. Both the heavy and light chain sequences exhibit a ratio of replacement to silent substitutions that range from 2:1 to 8:1 and fall mainly within the CDRs. There are, however, a significant number of replacement substitutions in FRs as well, an observation that has been documented previously, suggesting either a direct or indirect role for these residues in antigen binding. On the other hand, silent substitutions may either be the result of random mutational events introduced by the somatic hypermutation machinery or may represent polymorphism in a subfamily of germline gene segments. Without germline gene information from the original donors of each of these cell lines, these possibilities cannot be precisely assessed.

Not only do these cell lines provide an opportunity to examine a part of the human alloantibody repertoire, they are a source of HLA-typing reagents. Although much of the HLA-typing is currently performed using PCR-based methods, a large portion still depends on serological methods and, therefore, a panel of reagents which can distinguish individual polymorphisms within a subgroup of molecules. There are a number of murine monoclonal antibodies that recognize human HLA molecules, but often they do not distinguish polymorphic determinants. Therefore it is necessary to obtain alloantisera. Unfortunately there are drawbacks to alloantisera as well, such as consistently low titers of antibody and the need to absorb with platelets to remove anti-class I antibodies. The development of human monoclonal antibodies eliminates some of these problems, but often others remain, like low production of antibody and genetic instability of the antibody-producing cell lines.

Despite the disadvantages associated with the use of alloantisera, the epitopes defined by most human anti-HLA monoclonal antibodies represent polymorphic structures reflecting the allelic variation within the HLA system. Therefore, human monoclonal antibodies may aid in understanding the molecular mechanism of processes in which HLA polymorphism plays a critical role, such as graft rejection, antigen presentation, and susceptibility to certain HLA-associated autoimmune diseases, including insulin-dependent diabetes, celiac disease and rheumatoid arthritis. They can also serve as useful HLA-typing reagents for organ transplantation and molecular mapping of the allo-HLA polymorphic epitopes that are involved in antibody and T lymphocyte recognition. Nucleotide sequences of a large number of HLA class I and class II alleles are now available which aid in the definition of such epitopes and their location in the three-dimensional structure of HLA molecules. Since certain class II molecules may predispose some individuals to the development of particular autoimmune diseases, human anti-HLA monoclonal antibodies could provide a means of identification of the most important functional epitopes associated with disease susceptibility. In addition, human anti-HLA monoclonals may have therapeutical application in autoimmune disorders, as has been successfully achieved in animal models (Adelman *et al.*, 1983).

Human immunoglobulin repertoire studies continue to expand and currently include the structural analysis of the V_H and V_L gene segments expressed in B cells that have been exposed to a large number of different antigens. In general, it is apparent that human antibodies, regardless of their specificity or disease of origin, can be derived from a substantial number of different germline gene segments which are (likely) subjected to somatic hypermutation and clonal selection. However, the issue of biased gene usage remains to be more accurately defined.

CONCLUSIONS

In conclusion, we have cloned and sequenced the heavy and light chain variable region gene segments utilized by six human monoclonal alloantibodies specific for various human class I and class II HLA molecules. Analysis of these expressed sequences revealed the predominant utilization of V_H3 gene segments (four of six) with one V_H1 and one V_H4. Three of the light chains are κ and three are λ. The light chain V segments were shown to be derived from several gene families including $V_\lambda2$, $V_\lambda3$, $V_\kappa1$ and $V_\kappa3$. We conclude that the distribution of V_H and V_L gene families reflects the expected level of expression based on the size of the family. In addition, these expressed sequences appear to be derived from a diverse array of germline gene segments, indicating that there is not a restriction at the level of individual gene segments. Third, there is minimal apparent somatic mutation, compared to previous studies describing antibodies to a variety of other exogenous antigens. However, the mutation that is present appears to have been antigen driven, as the ratio of replacement to silent substitutions ranges from 2:1 to 8:1.

Acknowledgements—The authors thank Shirley Hall for her technical expertise in the operation of the ABI automated sequencer and the blood donors of the Immunohematologic Research Center AVIS of Bergamo, Italy. J. D. Capra holds the Edwin L. Cox Distinguished Chair in Immunology and Genetics at the University of Texas Southwestern Medical Center and is also supported in part by National Insitutes of Health AI-12127 and CA44016 and U. S. Army Research Office Prime Grant DAAL03-92-G-0215. J. S. Andris is supported in part by the Texas Department Ladies Auxiliary, Veterans of Foreign Wars and the Division of Cell and Molecular Biology National Institutes of Health Training Grant 5T32 GMO8203-05. M. P. Pistillo and G. B. Ferrara are supported in part by CNR Genetic Engineering 94.00056 PF99, Biotechnology and Bioinstrumentation 93.01108 PF 70 and AIRC 1994.

REFERENCES

Adelman N. E., Watling D. L. and McDevitt H. O. (1983) Treatment of (NZB × NZW)F_1 disease with anti-IA monoclonal antibodies. *J. exp. Med.* **158**, 1350–1355.

Anderson M. L. M., Szajnert M. F., Kaplan J. C., McColl L. and Young B. D. (1984) The isolation of a human Ig V_λ gene from a recombinant library of chromosome 22 and estimation of its copy number. *Nucleic Acids Res.* **12**, 6647–6661.

Andris J. S., Johnson S., Zolla-Pazner S. and Capra J. D. (1991) Molecular characterization of five human anti-human immunodeficiency virus type 1 antibody heavy chains reveals extensive somatic mutation typical of an antigen-driven immune response. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7783–7787.

Andris J. S., Ehrlich P. H., Ostberg L. and Capra J. D. (1992) Probing the human antibody repertoire to exogenous antigens: characterization of the H and L chain V region gene segments from anti-Hepatitis B virus antibodies. *J. Immun.* **149**, 4053–4059.

Andris J. S., Brodeur B. R. and Capra J. D. (1993) Molecular characterization of human antibodies to bacterial antigens: utilization of the less frequently expressed V_H2 and V_H6 heavy chain variable region gene families. *Molec. Immun.* **30**, 1601–1616.

Ausubel F., Brent R., Kingston R., Moore D., Seidman J., Smith J. and Struhl K. (1989) *Current Protocols in Molecular Biology* (Edited by Ausubel F., Brent R., Kingston R., Moore D., Seidman J., Smith J. and Struhl K.). Green Publishing Associates and Wiley–Interscience, New York.

Ayala-Avila J., Vazques J., Danielsson L., Fernandez de Cossio M. E. and Borrebaeck C. A. K. (1993) Sequence determination of variable region genes of two human monoclonal antibodies against *Neisseria meningitidis*. *Gene* **127**, 273–274.

Barbas III C. F., Crowe Jr J. E., Cababa D., Jones T. M., Zebedee S. L., Murphy B. R., Chanock R. M. and Burton D. R. (1992) Human monoclonal Fab fragments derived from a combinatorial library bind to respiratory syncytial virus F glycoprotein and neutralize infectivity. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10164–10168.

Barbas III C. F., Collet T. A., Amberg W., Roben P., Binley J. M., Hoekstra D., Cababa D., Jones T. M., Williamson R. A., Pilkington G. R., Haigwood N. L., Cabezas E., Satterthwait A. C., Sanz I. and Burton D. R. (1993) Molecular profile of an antibody response to HIV-1 as probed by combinatorial libraries. *J. molec. Biol.* **230**, 812–823.

Barrett D. J., Goodenow M. M., Harville T. O., Sleasman J. W. and White G. B. (1992) Ig H chain usage in human Ab to pneumococcal polysaccharide. *Pediat. Res.* **31**, 147A.

Berman J. E., Mellis S. J., Pollock R., Smith C. L., Suh H., Heinke B., Kowal C., Surti U., Chess L., Cantor C. R. and Alt F. W. (1988) Content and organization of the human Ig V_H locus: definition of three new V_H families and linkage to the Ig C_H locus. *Eur. molec. Biol. Org. J.* **7**, 727–738.

Bogard Jr W. C., Siegel S. A., Leone A. O., Damiano E., Shealy D. J., Ely T. M., Frederick B., Mascelli M. A., Siegel R. C., Machielse B., Naveh D., Kaplan P. M. and Daddona P. E. (1993) Human monoclonal antibody HA-1A binds to endotoxin via an epitope in the lipid A domain of lipopolysaccharide. *J. Immun.* **150**, 4438–4449.

Burton D. R., Barbas III C. F., Persson M. A. A., Koenig S., Chanock R. M. and Lerner R. A. (1991) A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10134–10137.

Capra J. D. and Kehoe J. M. (1975) Hypervariable regions, idiotypy, and the antibody combining site. *Adv. Immun.* **20**, 1–40.

Chomczynski P. and Sacchi N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analyt. Biochem.* **162**, 156–159

Chothia C., Lesk A. M., Gherardi E., Tomlinson I. M., Walter G., Marks J. D., Llewelyn M. B. and Winter G. (1992) Structural repertoire of the human V_H segments. *J. molec. Biol.* **227**, 799–817.

Chuchana P., Blancher A., Brockly F., Alexandre D., Lefranc G. and Lefranc M. (1990) Definition of the human immunoglobulin variable λ (IGLV) gene subgroups. *Eur. J. Immun.* **20**, 1317–1325.

Combriato G. and Klobeck H.-G. (1991) V_λ and J_λ–C_λ gene segments of the human immunoglobulin λ light chain locus are separated by 14 kb and rearrange by a deletion mechanism. *Eur. J. Immun.* **21**, 1513–1522.

Cook G. P., Tomlinson I. M., Walter G., Riethman H., Carter N. P., Buluwela L., Winter G. and Rabbitts T. H. (1994) A complete map of the human immunoglobulin V_H locus on the telomeric region of chromosome 14q. *Nature Genet.* **7**, 162–168.

Cuisinier A. M., Guigou V., Boubli L., Fougereau M. and Tonnelle C. (1989) Preferential expression of V_H5 and V_H6 immunoglobulin genes in early human B cell ontogeny. *Scand. J. Immun.* **30**, 492–497.

Dersimonian H., Long A., Rubenstein D., Stollar B. D. and Schwartz R. S. (1990) V_H genes in human autoantibodies. *Int. Rev. Immun.* **5**, 253–264.

Felgenhauer M., Koh J. and Ruker F. (1990) Nucleotide sequences of the cDNAs encoding the V-regions of H-and L-chains of a human monoclonal antibody specific to HIV-1-gp41. *Nucleic Acids Res.* **18**, 4927.

Gillies S. D., Dorai H., Wesolowski J., Majeau G., Young D., Boyd J., Gardner J. and James K. (1989) Expression of human anti-tetanus toxoid antibody in transfected murine myeloma cells. *Bio/Technol.* **7**, 799–804.

Grillot-Courvalin C., Brouet J.-C., Piller F., Rassenti L. Z., Lahaume S., Silverman G. J., Silberstein L. and Kipps T. J. (1992) An anti-B cell autoantibody from Wiskott–Aldrich syndrome which recognizes i blood group specificity on normal human B cells. *Eur. J. Immun.* **22**, 1781–1788.

Huber C., Schable H. F., Huber E., Klein R., Meindl A., Thiebe R., Lamm R. and Zachau H. G. (1993) The V_κ genes of the L regions and the repertoire of V_κ genes. *Eur. J. Immun.* **23**, 2868–2875.

Hughes-Jones N. C., Bye J. M., Beale D. and Coadwell J. (1991) Nucleotide sequences and three-dimensional modelling of the V_H and V_L domains of two human monoclonal antibodies specific for the d antigen of the human Rh-blood-group system. *Biochem. J.* **268**, 135–140.

Ikematsu H., Harindranath N., Euki Y., Notkins A. L. and Casali P. (1993) Clonal analysis of a human antibody response II: sequences of the V_H genes of human IgM, IgG, and IgA to rabies virus reveal preferential utilization of V_HIII segments and somatic hypermutation. *J. Immun.* **150**, 1325–1337.

Insel R. A., Adderson E. E. and Carroll W. L. (1992) The repertoire of human antibodies to the Haemophilus influenzae type b capsular polysaccharide. *Int. Rev. Immun.* **9**, 25–43.

Irigoyen M., Manheimer-Lory A., Gaynor B. and Diamond B. (1994) Molecular analysis of the human immunoglobulin V_HII gene family. *J. Clin. Invest.* **94**, 532–538.

Kedrierski W. and Porter J. (1991) A novel nonenzymatic procedure for removing DNA template from RNA transcription mixtures. *Biotechniques* **10**, 210–214

Klohe E., Pistillo M. P., Ferrara G. B., Goeken N. E., Greazel N. S. and Karr R. W. (1992) Critical role of HLA-DRb1 residue 58 in multiple polymorphic epitopes recognized by xenogeneic and allogeneic antibodies. *Hum. Immun.* **35**, 18–28.

Kuppers R., Zhao M., Hansmann M. L. and Rajewsky K. (1993) Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. *Eur. molec. Biol. Org. J.* **12**, 4955–4967.

Lafaille J. J., de Cloux A., Bonneville M., Takagaki Y. and Tonegawa S. (1989) Junctional sequences of T cell receptor gamma–delta genes: implications for gamma–delta T cell lineages and for a novel intermediate of V–(D)–J joining. *Cell* **59**, 859–870.

Larrick J. W., Danielsson L., Brenner C. A., Abrahamson M., Fry K. E. and Borrebaeck C. A. K. (1989*a*) Rapid cloning of rearranged immunoglobulin genes from human hybridoma cells using mixed primers and the polymerase chain reaction. *Biochem. biophys. Res. Commun.* **160**, 1250–1256.

Larrick J. W., Danielsson L., Brenner C. A., Wallace E. F., Abrahamson M., Fry K. E. and Borrebaeck C. A. K. (1989*b*) Polymerase chain reaction using mixed primers: cloning of human monoclonal antibody variable region genes from single hybridoma cells. *Bio/Technol.* **7**, 934–938.

Leoni J., Ghiso F., Goni F. and Frangione B. (1991) The primary structure of the Fab fragment of protein KAU, a monoclonal immunoglobulin M cold agglutinin. *J. biol. Chem.* **266**, 2836–2842.

Lewis A. P., Lemon S. M., Barber K. A., Murphy P., Parry N. R., Peakman T. C., Sims M. J., Worden J. and Crowe J. S. (1993) Rescue, expression, and analysis of a neutralizing human anti-hepatitis A virus monoclonal antibody. *J. Immun.* **151**, 2829–2838.

Liu M.-F., Robbins D. L., Crowley J. J., Sinha S., Kozin F., Kipps T. J., Carson D. A. and Chen P. P. (1989) Characterization of four homologous L chain variable region genes that are related to 6B6.6 idiotype positive human rheumatoid factor L chains. *J. Immun.* **142**, 688–694.

Marasco W. A., Bagley J., Zani C., Posner M., Cavacini L., Haseltine W. A. and Sodroski J. (1990) Characterization of the cDNA of a broadly reactive neutralizing human anti-gp120 monoclonal antibody. *J. clin. Invest.* **90**, 1467–1478.

Matsuda F., Shin E. K., Nagaoka H., Matsumura R., Haino M., Fukita Y., Taka-ishi S., Imai T., Riley J. H., Anand R., Soeda E. and Honjo T. (1993) Structure and physical map of 64 variable segments in the 3′ 0 8 megabase region of the human immunoglobulin heavy-chain locus. *Nature Genet.* **3**, 88–94.

Mazzoleni O., Longo A., Angelini G., Colonna M., Tanikaki N., Delfino L., Pistillo M. P., Kun L. and Ferrara G. B. (1989) Human monoclonal antibody MP8 detects a supertypic determinant encoded by DPB alleles DPB2.1, DPB3, DPB4.2, DPB8, DPB9, DPB10, and DPB14. *Immunogenetics* **30**, 502–505.

Mazzoleni O., Pistillo M. P., Falco M., Tazzari P. L. and Ferrara G. B. (1991) Characterization of two human monoclonal antibodies recognizing HLA-A30 and HLA-A3+ A31, respectively. *Tissue Antigens* **38**, 224–227.

Meek K. D., Hasemann C. A. and Capra J. D. (1989) Novel rearrangements at the immunoglobulin D locus: inversions and fusions add to the IgH somatic diversity. *J. exp. Med.* **170**, 39–57.

Moran M. J., Andris J. S., Matsumato Y., Capra J. D. and Hersh E. M. (1993) Variable region genes of anti-HIV human monoclonal antibodies: non-restricted use of the V gene repertoire and extensive somatic mutation. *Molec. Immun.* **30**, 1543–1551.

Mortari F., Newton J. A., Wang J. Y. and Schroeder Jr H. W. (1992) The human cord blood antibody repertoire: frequent usage of the V_H7 gene family. *Eur. J. Immun.* **22**, 241–245.

Mortari F., Wang J. Y. and Schroeder Jr H. W. (1993) Human cord blood antibody repertoire: mixed population of V_H gene

segments and CDR3 distribution in the expressed C_κ and C_γ repertoires. *J. Immun.* **150**, 1348–1357.

Newkirk M. M., Gram H., Heinrich G. F., Ostberg L. and Capra J. D. (1988) Complete protein sequences of the variable regions of the cloned heavy and light chains of a human anti-cytomegalovirus antibody reveal a striking similarity to human monoclonal rheumatoid factors of the Wa idiotypic family. *J. clin. Invest.* **81**, 1511–1518.

Pargent W., Meindl A., Theibe R., Mitzel S. and Zachau H. G. (1991) The human immunoglobulin κ locus: characterization of the duplicated O regions. *Eur. J. Immun.* **21**, 1821–1827.

Pascual V., Randen I., Thompson K., Sioud M., Forre O., Natvig J. and Capra J. D. (1990) The complete nucleotide sequences of the heavy chain variable regions of six monospecific rheumatoid factors derived from Epstein–Barr virus-transformed B cells isolated from the synovial tissue of patients with rheumatoid arthritis. *J. clin. Invest.* **86**, 1320–1328.

Pascual V., Victor K., Lelsz D., Spellerberg M. B., Hamblin T. J., Thompson K. M., Randen I., Natvig J. B., Capra J. D. and Stevenson F. K. (1991) Nucleotide sequence analysis of the V regions of two IgM cold agglutinins: evidence that the V_H4–21 gene segment is responsible for the major cross-reactive idiotype. *J. Immun.* **146**, 4385–4391.

Pascual V., Victor K., Randen I., Thompson K., Steinitz M., Forre O., Fu S.-M., Natvig J. B. and Capra J. D. (1992*a*) Nucleotide sequence analysis of rheumatoid factors and polyreactive antibodies derived from patients with rheumatoid arthritis reveals diverse use of V_H and V_L gene segments and extensive variability in CDR3. *Scand. J. Immun.* **36**, 349–362.

Pascual V., Victor K., Spellerberg M., Hamblin T. J., Stevenson F. K. and Capra J. D. (1992*b*) V_H restriction among human cold agglutinins: the V_H4–21 gene segment is required to encode anti-I and anti-i specificities. *J. Immun.* **149**, 2337–2344.

Pascual V., Verkruyse L., Casey M. L. and Capra J. D. (1993) Analysis of Ig H chain gene segment utilization in human fetal liver. *J. Immun.* **151**, 4164–4172.

Pascual V., Liu Y., Magalski A., de Bouteiller O., Banchereau J. and Capra J. D. (1994) Analysis of somatic mutation in five B cell subsets of human tonsil. *J. exp. Med.* **180**, 329–339.

Pascual V. and Capra J. D. (1990) Human immunoglobulin heavy chain variable region genes: organization, polymorphism, and expression. *Adv. Immun.* **49**, 1–74.

Pascual V. and Capra J. D. (1992) V_H4-21, a human gene segment overrepresented in the autoimmune repertoire. *Arthrit. Rheum.* **35**, 11–18.

Perlmutter R. M., Kearney J. F., Chang S. P. and Hood L. E. (1985) Developmentally controlled expression of immunoglobulin V_H genes. *Science* **227**, 1597–1601.

Persson M. A. A., Caothien R. H. and Burton D. R. (1991) Generation of diverse high affinity human monoclonal antibodies by repertoire cloning. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2432–2436.

Pistillo M. P., Hammerling U., Dupont B. and Ferrara G. B. (1986) *In vitro* production of a human HLA alloantibody of restricted specificity (DQw2) via Epstein–Barr virus transformation. *Hum. Immun.* **15**, 109–117.

Pistillo M. P., Tanigaki N., Mazzoleni O., Ciccone E., Hammerling U., Park M., Terasaki P. I. and Ferrara G. B. (1987) Human lymphoblastoid cell lines secreting antibodies with restricted HLA specificity. *Immunogenetics* **25**, 145–151.

Pistillo M. P., Mazzoleni O., Tanigaki N., Hammerling U., Longo A., Frumento G. and Ferrara G. B. (1988) Human anti-HLA monoclonal antibodies: production, characterization, and application. *Hum. Immun.* **21**, 265–278.

Pistillo M. P., Tanigaki N., Chua R., Mazzoleni O. and Ferrara G. B. (1989) Human anti-HLA-DQw2 monoclonal antibody secreted by an Epstein–Barr virus-transformed lymphoblastoid cell line: assessment of the monoclonality, allospecificity, and target. *Hum. Immun.* **24**, 253–263.

Pistillo M. P., Mazzoleni O., Kun L., Falco M., Tazzari P. L. and Ferrara G. B. (1991) Production of two human hybridomas secreting antibodies to HLA-DRw11 and -DRw8+ w12 specificities. *Hum. Immun.* **31**, 86–93.

Pistillo M. P., Sun P. F., Mantero S. and Ferrara G. B. (1993) Shared epitopes of the HLA-DR10 molecule recognized by murine and human mAbs. *Eur. J. Immunogenet.* **20**, 111–121.

Pistillo M. P., Bini D., Pozzi S. and Ferrara G. B. (1995) A cytotoxic human monoclonal antibody that can discriminate HLA-A*3002 from HLA-A*3001 subtype. *Tissue Antigens* (in press).

Rioux J. D., Larose Y., Brodeur B. R., Radzioch D. and Newkirk M. M. (1994) Structural characteristics of four human hybridoma antibodies specific for the pp65 protein of the human cytomegalovirus and their relationship to human rheumatoid factors. *Molec. Immun.* **31**, 585–597.

Rudikoff S., Pawlita M., Pumphrey J., Mushinski E. and Potter M. (1983) Galactan-binding antibodies: diversity and structure of idiotypes. *J. exp. Med.* **158**, 1385–1400.

Sanger F., Nicklen S. and Coulson A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.

Sanz I., Kelly P., Williams C., Scholl S., Tucker P. and Capra J. D. (1989) The smaller human V_H gene families display remarkable little polymorphism. *Eur. molec. Biol. Org. J.* **8**, 3741–3748.

Sanz I. (1991) Multiple mechanisms participate in the generation of diversity of human H chain CDR3 V regions. *J. Immun.* **147**, 1720–1729.

Schilling J., Clevinger B., Davie J. M. and Hood L. (1980) Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V-region gene segments. *Nature* **283**, 35–40.

Schroeder Jr H. W., Hillson J. L. and Perlmutter R. L. (1987) Early restriction of the human antibody repertoire. *Science* **238**, 791–793.

Schroeder Jr H. W., Stuber F. S., Kirkham P.M., Gray B. M., Tzeng J. J. and Shaw D. R. (1992) Sequence analysis of two human monoclonal antibodies with specificity for type 3 pneumococcal polysaccharides (PPS-3). *FASEB J.* **6**, A1223.

Schroeder Jr H. W. and Wang J. Y. (1990) Preferential utilization of conserved immunoglobulin heavy chain variable gene segments during human fetal life. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6146–6150.

Scott M. G., Zachau H. G. and Nahm M. H. (1992) The human antibody V region repertoire to the type b capsular polysaccharide of *Haemophilus influenzae*. *Int. Rev. Immun.* **9**, 45–55.

Silberstein L. E., Jefferies L. C., Goldman J., Friedman D., Moore J. S., Nowell P. C., Roelcke D., Pruzanski W., Roudier J. and Silverman G. J. (1991) Variable region gene analysis of pathogenic human autoantibodies to the related i and I red blood cell antigens. *Blood* **9**, 2372–2386.

Songsivilai S., Bye J. M., Marks J. D. and Hughes-Jones N. C. (1990) Cloning and sequencing of human λ immunoglobulin genes by the polymerase chain reaction. *Eur. J. Immun.* **20**, 2661–2666.

Stevenson F. K., Wrightham M., Glennie M. J., Jones D. B., Cattan R. R., Feizi T., Hamblin T. J. and Stevenson G. T.

(1986) Antibodies to shared idiotypes as agents for analysis and therapy for human B cell tumors. *Blood* **68**, 430–436.

Stiernholm N. B. J., Kuzniar B. and Berinstein N. L. (1994) Identification of a new human V-λ gene family–V-λ X. *J Immun.* **152**, 4969–4975.

Straubinger B., Huber E., Lorenz W., Osterholzer E., Pargent W., Pech M., Pohlenz H. D., Zimmer F. J. and Zachau H. G. (1988) The human VK locus: characterization of a duplicated region encoding 28 different immunoglobulin genes. *J. molec. Biol.* **199**, 23–34.

Thompson K. M., Sutherland J., Barden G., Melamed M. D., Randen I. and Natvig J. B. (1991) Human monoclonal antibodies against blood group antigens preferentially express a V_H4-21 variable region gene-associated epitope. *Scand. J. Immun.* **34**, 509–518.

Tonegawa S. (1983) Somatic generation of antibody diversity. *Nature* **302**, 575–581.

Tuaillon N., Taylor L. D., Lonberg N., Tucker P. W. and Capra J. D. (1993) Human immunoglobulin heavy-chain minilocus recombination in transgenic mice: gene-segment use in mu and gamma transcripts. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3720–3724.

van Dijk K. W., Mortari F., Kirkham P. M., Schroeder Jr H. W. and Milner E. C. B. (1993) The human immunoglobulin V_H7 gene family consists of a small polymorphic group of six to eight gene segments dispersed throughout the V_H locus. *Eur. J. Immun.* 832–839.

van Es J., Meyling F. H. J. G., van de Akker W., Aanstoot H., Derksin R. H. W. M. and Logtenberg T. (1991) Somatic mutations in the variable regions of a human IgG anti-double-stranded DNA autoantibody suggest a role for antigen in the induction of systemic lupus erythematosus. *J. exp. Med.* **173**, 461–470.

Victor K. and Capra J. D. (1994) An apparently common mechanism of generating antibody diversity: length variation of the V_L–J_L junction. *Molec. Immun.* **31**, 39–46.

Williams S. C. and Winter G. (1993) Cloning and sequencing of human immunoglobulin variable λ gene segments. *Eur. J. Immun.* **23**, 1456–1461.

Yamada M., Wasserman R., Reichard B. A., Shane S., Caton A. J. and Rovera G. (1991) Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. *J. exp. Med.* **173**, 395–407.

Yancoupoulos G. D., Desiderio S. V., Paskind M., Kearney J. F., Baltimore D. and Alt F. W. (1984) Preferential utilization of the most J_H-proximal V_H gene segments in pre-B cell lines. *Nature* **311**, 727–733.

Zachau H. G. (1989) Immunoglobulin light chain genes of the κ type in man and mouse. In *Immunoglobulin Genes* (Edited by Honjo T., Alt F. W. and Rabbitts T. H.), p. 91. Academic Press, New York.

Zebedee S. L., Barbas III C. F., Hom Y. L., Caothien R. H., Graff R., DeGraw J., Pyati J., LaPolla R., Burton D. R., Lerner R. A. and Thornton G. B. (1992) Human combinatorial antibody libraries to hepatitis B surface antigen. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3175–3179

Exhibit M

Scand. J. Immunol. 42, 442–448, 1995

V-Region and Class Specific RT-PCR Amplification of Human Immunoglobulin Heavy and Light Chain Genes from B-Cell Lines

B. A. WATKINS, A. E. DAVIS, S. FIORENTINI & M. S. REITZ JR

Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

Watkins BA, Davis AE, Fiorentini S, Reitz Jr MS. V-Region and Class Specific RT-PCR Amplification of Human Immunoglobulin Heavy and Light Chain Genes from B-Cell Lines. Scand J Immunol 1995; 42:442–448

We have designed and tested primers that amplify complete human kappa and lambda light chain genes, and human Fd fragments from gamma, mu and alpha heavy chain genes. These primers were tested for efficiency and specificity on monoclonal sources of human immunoglobulin RNA, obtained from human B-cell lines of known immunoglobulin gene expression. Analysis of the sequences derived from these B-cells confirms the specificity of the PCR primers and the extent of somatic mutation seen in different B-cell malignancies supports existing concepts for differing aetiologies in the tumours concerned.

B.A. Watkins, Laboratory of Tumor Cell Biology, National Cancer Institute, Room 6A11, 37 Convent Drive, MSC 4255, Bethesda, MD 20892-4255, USA

INTRODUCTION

The production of combinatorial antibody libraries in bacteria is based on the efficient cloning of active immunoglobulin genes into bacterial expression vectors. This is usually achieved by using the polymerase chain reaction (PCR) to amplify complementary DNA (cDNA) of the active immunoglobulin genes with primers containing restriction sites which enables directional cloning into bacterial expression vectors [1–5]. The complexity and representation of the library achieved depends largely on the ability of the primers used to amplify a broad spectrum of immunoglobulin variable region genes. Increased representation within a library can also be achieved by increasing the number of classes of immunoglobulin genes included. In designing primers for repertoire cloning, we decided to examine the specificity of PCR primers used for the amplification of human active human immunoglobulin V-region genes, and to extend the number of immunoglobulin classes included. In order to be compatible with existing Fab phage display vectors [2, 3], primers were designed to amplify the major V_κ and V_λ gene families as intact light chains, and to amplify V_H gene families as Fd fragments from IgG, IgM and IgA. To test these primers we amplified and cloned several light and heavy chain immunoglobulin genes from clonal populations of B-cells. As a source of clonal B-cell populations to test our primers, we selected B-cell lines from a Burkitt's lymphoma (Daudi) [6], two multiple myelomas (RPMI 8226 [7] and IM-9 [8]), an undifferentiated B-cell lymphoma (MC116) [9] and *in vitro* Epstein–Barr Virus (EBV) transformed B-cells (Dakiki) [10].

MATERIALS AND METHODS

Cell culture. Cell lines RPMI 8226 (CCL 155), IM-9 (CCL 159), Daudi (CCL 213), Dakiki (TIB 206) and MC116 (CRL 1649) were obtained from ATCC (Rockville, MD, USA) and grown in RPMI 1640, supplemented with 25 μg/ml Gentamicin, 2 mM L-Glutamine and 20% FBS. Each cell line was grown until the number of cells exceeded 10^8, at which point they were harvested, washed once in PBS, and then used for the preparation of RNA.

RNA extraction and cDNA synthesis. Total cellular RNA was extracted using standard techniques [11]. Briefly, 10^8–10^9 cells from each cell line were dissolved by homogenization in 10 ml 4M guanidine thiocyanate supplemented with 0.1% sarcosine and 80 μl beta-mercaptoethanol. DNA was then sheared by 10 passes through a 21 gauge needle, followed by 10 passes through a 25 gauge needle. This cellular homogenate was then loaded onto a cesium chloride cushion (density = 1.62 g/ml), and centrifuged at 300 000 *g* for 16 h/20°C in a Beckman Ti50 rotor. Total RNA was recovered as a pellet, washed with ethanol, resuspended in 500 μl DEPC-treated water, precipitated with 50 μl of 3M sodium acetate and 1 ml of ethanol, resuspended in a final volume of 300 μl and quantified by absorbance at 260 nm. Typical yields were 1–4 μg/μl for a total of 300 μg to 1.2 mg from 10^8 cells. 5–25 μg of total RNA was denatured by incubation at room temperature (23°C) for 5 min with 1 mM methyl

Table 1. Primers used for PCR amplification of human immunoglobulin genes

Chain primer		Sequence	Specificity
	1a	AGA TGT GAGCTC CAG ATG ACC CAG TCT CC	$V_{\kappa 1}$&$V_{\kappa 4}$
	1b	CAG TGG GAGCTC GTG ATG ACT CAG TCT CC	$V_{\kappa 2}$&$V_{\kappa 6}$
	1c	ACC GGA GAGCTC GTG TTG ACG CAG TCT CC	$V_{\kappa 3}$&$V_{\kappa 5}$
	2	GCG CCG TCTAGA ACT AAC ACT CTC CCC TGT TGA AGC TCT TTG TGA CGG GCG AAC TCA G	C_κ
Light	3a	GCG ATC GAGCTC TCT GTG CTG ACT CAG CC	$V_{\lambda 1}$
	3b	TCC TGG GAGCTC TCT GCC CTG ACT CAG CC	$V_{\lambda 2}$
	3c	TCT GTG GAGCTC TAT GTG CTG ACT CAG CC	$V_{\lambda 3}$
	3d	TCT GTG GAGCTC TCT GAG CTG ACT CAG GA	$V_{\lambda 4}$
	3e	TCC AAT GAGCTC ACT GTG GTG ACT CAG GA	$V_{\lambda 7}$
	4	GCG CCG TCTAGA CTA AGA ACA TTC TGC AGG GGC CA	C_λ
	5a	G GTC CTG CTCGAG GTG CAG CTG GTG CAG TCT GG	V_{H1}&V_{H5}
	5b	G GTC CTG CTCGAG GTG CAG CTG CAG GAG TCG GG	V_{H4}
	5c	GTC CTG CTCGAG GTG CAG CTG GTG GAG TCT GG	V_{H3}
	5d	GTC CTG CTCGAG GTC ACC TTG AAG GAG TCT GG	V_{H2}
	5e	G GTC CTG CTCGAG GTG CAG CTA CAG CAG TGG GG	V_{H4-21}
Heavy	5f	T GTC CTG CTCGAG GTA CAG CTG CAG CAG TCA GG	V_{H6}
	6a	CAG AGT ACTAGT CTT GTC CAC CTT GGT GTT GCT	IgG
	6b	GTG AGT ACTAGT ACA AGA TTT GGG CTC AAC T	IgG_1
	6c	CTC AGC ACTAGT TGG TAG AGG CAC GTT CTT TT	IgM
	6d	CAG AGT ACTAGT TGG GCA GGG CAC AGT CAC AT	IgA

Restriction sites inserted for cloning are underlined.

mercury hydroxide. The RNA was then put on ice, mixed with the remaining reagents, then incubated for 1 h at 42°C. The reaction conditions were: 35 mM beta-mercaptoethanol; 1 mM each dNTP; 0.1 μg/μl random hexamers; 10 mM Tris HCl pH 8.8; 50 mM KCl; 6.5 mM $MgCl_2$; 0.01% Gelatin; 500 units/mL RNasin; 2 units/μl AMV RT. cDNA was used directly for PCR.

Polymerase chain reaction. Reaction volumes were 100 or 200 μl using 5 or 10 μl of a cDNA reaction for each PCR reaction. Conditions for PCR were 10 mM Tris-HCl pH 8.8; 50 mM KCl; 1.5 mM $MgCl_2$; 0.01% gelatin, 160 μM each dNTP, 2.5 μM each primer. Reaction parameters were 94°C/1 min; 45°C/2 min; 72°C/1 min, 40 cycles. The final cycle was followed by a 7 min/72°C primer extension phase.

The primers used for the PCR amplification of human immunoglobulin genes are shown in Table 1. Primers 1a ($V_{\kappa 1}$ and $V_{\kappa 4}$), 1b ($V_{\kappa 2}$ and $V_{\kappa 6}$) or 1c ($V_{\kappa 3}$ and $V_{\kappa 5}$) were used with primer 2 in separate reactions for the amplification of full length kappa chains. Primer 2 was designed to incorporate a C to T mutation to eliminate a naturally occurring Sac-I site in human C_κ gene. Primers 3a($V_{\lambda 1}$), 3b($V_{\lambda 2}$), 3c($V_{\lambda 3}$), 3d($V_{\lambda 4}$) and 3e($V_{\lambda 7}$) were used with primer 4 to amplify full-length lambda immunoglobulin light chains.

Amplification of heavy chain Fd fragments was achieved using primers 5a through 5f with primers 6a (γ_{1-4}), 6b (γ_1), 6c (μ) or 6d (α_1 and α_2). Primer 5a was designed to recognize group I heavy chain variable (V_{H1}) domains, but due to the homology between groups V_{H1} and V_{H5}, should also recognize V_{H5} genes. Primer 5b was designed to amplify V_{H2} genes and primer 5c to recognize V_{H3} genes. Primers 5d and 5e were designed to amplify the V_{H4} family of genes, primer 5e being designed to recognize the V_{H4-21} group of germline sequences. Primer 5f was designed to recognize sequences derived from the single V_{H6} germline sequence.

Cloning. PCR products were separated by electrophoresis through a 1.5% agarose gel. Bands at approximately 650 bp were isolated using with DEAE paper [11], then digested with the appropriate restriction enzymes and cloned into pBluescript-II KS- (Stratagene, La Jolla, CA, USA) using standard techniques [11]. Clones containing inserts were sequenced using standard dideoxy techniques [11]. The heavy chain Fd fragment derived from cell line Dakiki was initially refractory to cloning. Post restriction digestion analysis of the PCR product revealed bands of approximately 500bp and 150bp. This Fd fragment was then cloned in two parts, as an Xbo-I/Spe-I fragment of 140bp, and as an Spe-I/Spe-I fragment of 510bp, demonstrating that the additional Spe-I site was in the V_H region, which was confirmed by sequencing.

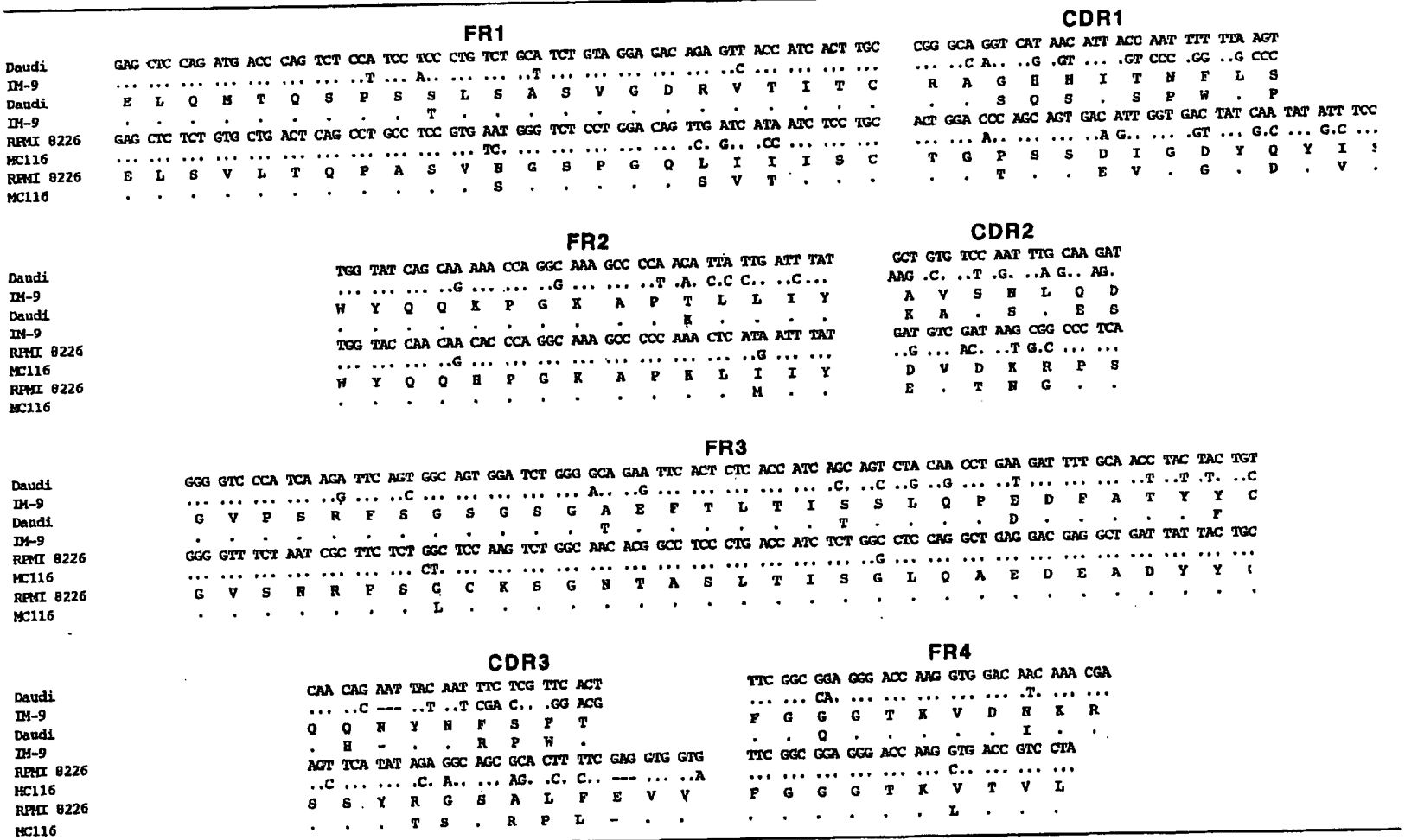
Database homology searches. The sequences described in this report were compared with the December 1994 update of genbank and EMBL databases. Genbank has assigned the following accession to the sequences described in this report: Cell line IM-9; IgG Fd: U07985; κ chain: U07989; Cell line RPMI 8226; λ Chain: U07992. Cell line Dakiki, IgA Fd: U07986. Cell line Daudi; IgM Fd: U07987, κ chain: U07990. Cell line MC116; IgM Fd: U07988, λ chain: U07991.

RESULTS

Immunoglobulin light chains

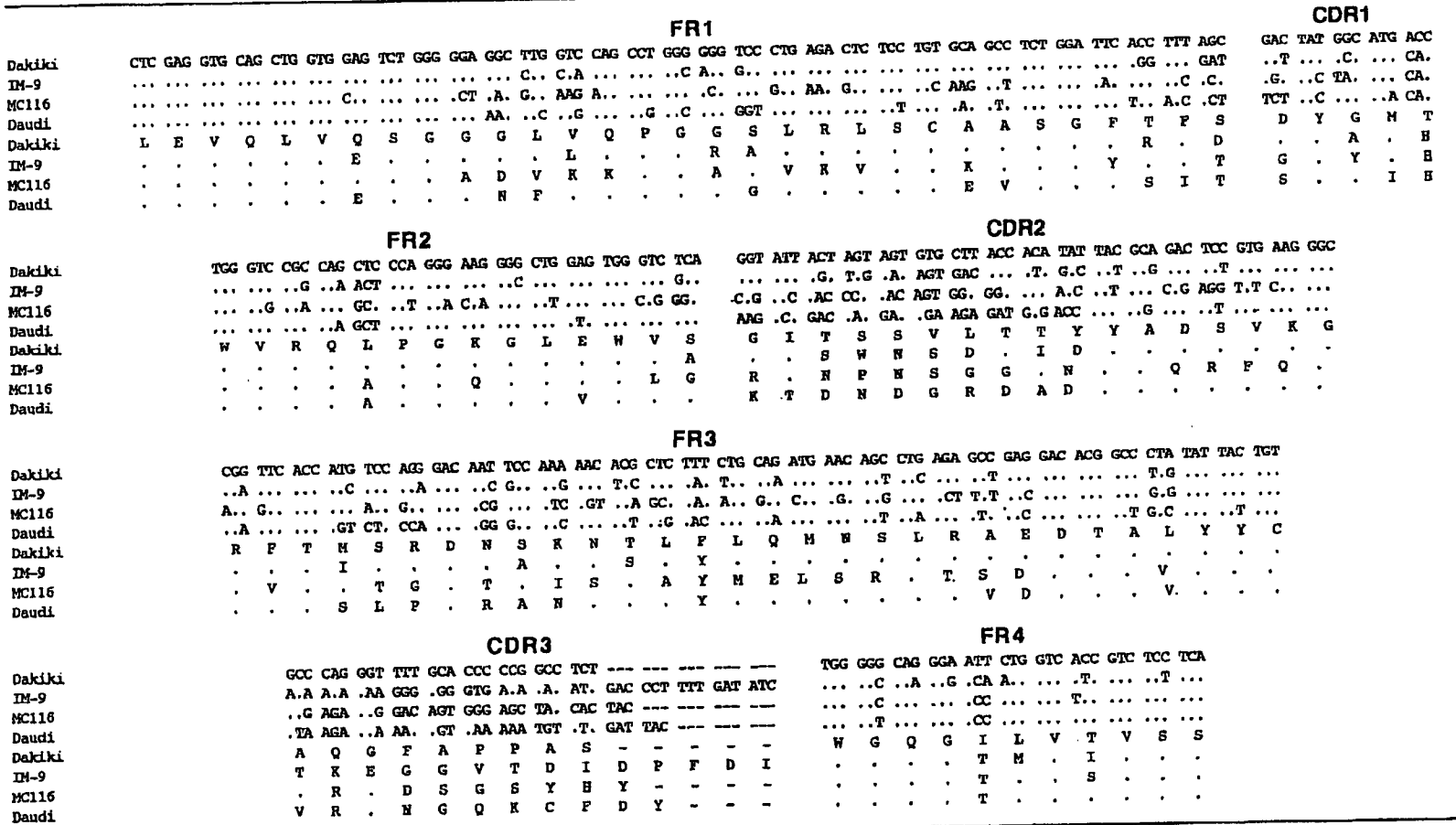
We were able to amplify four active light chain genes from the 5 B-cell lines tested, the sequences of which are shown in Table 2. Kappa light chains were amplified from cell lines Daudi and IM-9 with the $V_{\kappa 1}$ primer alone, and sequencing

Table 2. Active immunoglobulin light chain V-J region genes from B-cell lines Daudi, IM-9, MC116 and RPMI 8226



Rows 1, 2, 5 and 6 of each stanza show the DNA sequences; rows 3, 4, 7 and 8 show the deduced amino acid sequences.
Dots indicate homology, dashes indicate bases or amino acids that are not represented in that sequence.

Table 3. Heavy chain VDJ-region sequences from B-cell lines Dakiki, IM-9, MC116 and Daudi



Upper 4 rows of each stanza show the DNA sequences, the lower 4 show the deduced amino-acid sequences.
Dots indicate homology, dashes indicate bases or amino acids that are not represented in that sequence.

confirmed that both V-region genes were from the $V_{\kappa I}$ family. PCR amplification of lambda light chains using cDNA from MC116 RNA as a template gave positive reactions for $V_{\lambda 1}$ and $V_{\lambda 3}$, with the reaction for $V_{\lambda 1}$ being considerably stronger than that for $V_{\lambda 3}$. Using cDNA from RPMI 8226 RNA as a template, positive reactions were obtained with primers for $V_{\lambda 1}$ and $V_{\lambda 2}$, with the reaction for $V_{\lambda 1}$ being stronger than that for $V_{\lambda 2}$. Sequence analysis of both sequences revealed that they belong to the $V_{\lambda 2}$ family. The greater amplification using a $V_{\lambda 1}$ primer was therefore surprising and we failed to amplify immunoglobulin light chain from cDNA derived from Dakiki RNA. A positive PCR signal was obtained with primers designed to amplify only the Cl region and a positive signal corresponding to a full length lambda transcript was obtained on a northern blot probed with a C_λ probe. Additional primers were synthesized corresponding to variants of $V_{\lambda 1}$ and $V_{\lambda 4}$, but these also failed to amplify the full length lambda transcript. We conclude from these observations that the lambda light chain expressed in cell line Dakiki is significantly different at the site of our 5′ V_λ PCR primers than any of the primers tested thus far.

The Daudi kappa chain contains a $J_{\kappa 4}$ sequence and the IM-9 light chain uses a $J_{\kappa 1}$ J-region. Cell lines MC-116 and RPMI8226 both utilize $J_{\lambda 2}$ derived J-regions and $C_{\lambda 2}$ constant regions.

Comparison of the nucleotide sequences showed that the sequence for the Daudi V_κ region differed from the previously reported sequence [12] at three nucleotides, one of which resulted in an amino acid change (a V to D change at a.a. 56). We have sequenced three clones of the Daudi V_κ from three separate cDNA syntheses; all three clones have the same sequence, indicating that the observed divergence from the previously reported sequence was real. It is not known whether these differences reflect sequencing error in the original report, or on-going somatic mutation in the cell line.

The Daudi $V_{\kappa 1}$ gene shows the greatest homology to germline gene L11 [13], with which it was only 84.6% identical. Comparison with rearranged $V_{\kappa 1}$ genes indicated that the Daudi $V_{\kappa 1}$ region shared less than 90% with all sequences on the database. The IM-9 $V_{\kappa 1}$ gene was 96.3% identical to germline $V_{\kappa 1}$ gene L12 [14], and 94.9% identical to germline gene $H_\kappa 102$ [15]. The MC116 V_λ region was highly homologous (89–96%) to many rearranged $V_{\lambda 2}$ genes, but was most closely related to one germline $V_{\lambda 2}$ sequence (DPL13) [16], with which it shared 96% homology.

The RPMI8226 $V_{\lambda 2}$ gene showed 91.4% identity to the same germline $V_{\lambda 2}$ gene to which the $V_{\lambda 2}$ gene cloned from cell line MC116 (DPL13) [16] was most closely related. Given the limited homology of the RPMI8226 $V_{\lambda 2}$ gene to DPL13, it is possible that this gene is related to a previously undescribed germline $V_{\lambda 2}$ gene. The RPMI8226 $V_{\lambda 2}$ gene was, however, completely identical to a rearranged $V_{\lambda 2}$ gene ($V_\lambda 001$) [17]. It is remarkable that two rearranged genes from separate B-cell lines should have identical $V_{\lambda 2}$-$J_{\lambda 2}$ sequences, particularly when the V_λ sequence concerned has diverged from the nearest known germline gene by almost 9% (25 altered nucleotides), possibly indicating cross-contamination of the two cell lines.

Immunoglobulin heavy chains

Four immunoglobulin heavy chain Fd fragments were successfully amplified and cloned, the sequences of which are shown in Table 3. An IgG_1 Fd fragment was cloned from IM-9, an IgA_1 Fd fragment was cloned from Dakiki, and IgM Fd fragments were cloned from cell lines Daudi and MC116. No attempt was made to amplify Fd fragments from cell line RPMI 8226 on the basis of previous reports of light chain expression only [7]. The active VH domains from cell lines Daudi, Dakiki and IM-9 are from the V_{H3} family, while that from cell line MC116 is from the V_{H1} family. Under the conditions employed, there was little cross-recognition of V_H regions by primers designed to amplify different V_H regions. Using cDNA derived from Dakiki, we obtained specific amplification with primers 5c and 6d (V_{H3}/IgA), using cDNA from IM-9, amplification was only observed with primers 5c and 6a (V_{H3}/IgG), and using cDNA derived from Daudi amplification was only seen with primers 5c and 6c (V_{H3}/IgM). Only in the case of MC116 was some reactivity seen with primers 5c and 6d on a V_{H1} template, in addition to amplification with primers 5a and 6d. Primers 6a and 6b both amplified an IgG1 Fd fragment from IM-9 derived cDNA with primer 5c. Both PCR products were cloned and sequenced, and both were identical, with the exception of the 3′ end where the clone obtained using primer 6b contained the N-terminal part of the IgG_1 hinge as expected. Analysis of J-gene usage showed that three of the four heavy chains cloned (Daudi, Dakiki and MC116) contained J_{H4} related J-regions, while the IM-9 heavy chain contained a J_{H3} related sequence.

The V_{H3} gene from cell line Daudi was most closely related to human germline gene DP53 [18]/H11 [19], showing 84% homology, but was almost as related (80–82% homologous) to germline genes DP87 [20], DP58 [18], HHG19 [21] and DP54 [18], as well as numerous rearranged genes. The V_{H3} gene cloned from IM-9 was 94.6% identical to germline gene DP31 [18], and not closely related to any other germline V_H gene. The Dakiki V_{H3} gene was 93% homologous to three germline genes DP47 [18], V-B19.7 [22] and VH26 [23], and was more than 90% homologous to many rearranged V_H gene sequences. The V_{H1} gene cloned from MC116 was closely related to several germline genes: it showed 97.3% and 96.9% identity to germline genes DP75 and DP8, respectively [18], and more than 90% identical to two other germline genes, as well as five rearranged genes.

DISCUSSION

We have designed oligonucleotide PCR primers which allow amplification of human kappa and lambda light chains, and

alpha, gamma and mu heavy chain Fd fragments. The kappa light chain primers were efficient in amplifying $V_{\kappa 1}$ containing kappa chain genes, and are similar to previously described primers [5, 24]. We were able to amplify two of the three lambda light chains we attempted, but not necessarily with the primers predicted, indicating that further work may be required to optimize the site and/or design of V_λ primers. Amplification of heavy chain Fd fragments was specific, both for class of heavy chain and for family of V_H region. We had tried without success to amplify the active heavy chain Fd fragments with previously described V_H region primers [5], and arrived at the present V_H region sequences by trial and error. The strategy used in the placement of the Xho-1 site, and the overall V_H region primer sequences are similar, although not identical, to recently published V_H region primers [25].

The primers we described have been designed to be compatible with previously described vectors for the display of Fab fragments on the surface of filamentous phage. While we have improved the primers available to work with these vectors and expanded the number available to include both IgM and IgA, there are some additional questions that have arisen. Specifically, the discovery of an internal Spe-I restriction site in one of four heavy chain variable region genes we cloned clearly indicates that any Spe-I site containing genes would clearly not be represented in an antibody library constructed using Spe-I. This, in turn, suggests that the use of Spe-1 for antibody library construction is less than ideal. An additional question exists regarding the feasibility of cloning of IgM Fd genes and their expression as Fab fragments, due to the lack of a simple hinge in IgM. While it has been demonstrated that such a molecule can be expressed using a flexible linker to join the heavy and light chains [26], it may also be possible to modify the IgM reverse primer (6c in Table 1) to encode the sequence PCP (as with the IgA primer 6d) in place of PLP, thereby allowing the Fd fragment to covalently link to the light chain. We have tested such a primer and found it to amplify IgM Fd fragments with the same efficiency as primer 6c.

From the immunoglobulin sequences that we obtained from these B-cell lines, there are clearly differences in the extent to which the different genes have varied from known germline genes. We were also able to compare the extent of variation of V_H and V_L genes from the same B-cell line for three of the cell lines we used. The genes from MC116 (derived from an undifferentiated lymphoma) were closest to germline sequences. The V_H gene was closely related (97.3% and 96.9% identical) to two V_{H1} germline genes, and the V_λ gene was 96% homologous to a germline V_λ gene. Similarly, the V-region genes from multiple myeloma-derived cell-line IM-9 were 94.6% (V_{H3}) and 96.3% ($V_{\kappa 1}$) homologous to the nearest germline genes. In contrast, the V region gene sequences from Burkitt's lymphoma derived cell line Daudi were highly divergent from both germline and rearranged gene sequences. Both V_{H3} and $V_{\kappa 1}$ genes were approximately 84% homologous to the nearest germline sequences, and almost equally distant from the most homologous rearranged genes. This lack of homology could indicate extensive somatic mutation, which would be unusual for a cell that had not undergone the class switching from IgM to IgG (or IgA) normally thought to be associated with somatic mutation of active immunoglobulin genes. A recent report has demonstrated ongoing intraclonal variation within a follicular lymphoma expressing an IgM immunoglobulin [27], thought to be due to ongoing somatic mutation driven by antigen. Although no such observation has been made in Burkitt's lymphoma, it is clearly possible for extensive somatic mutation to occur in the absence of class switching. A possible alternative explanation is that the germline genes from which the Daudi sequences were derived have not yet been characterized.

Overall, the sequences derived from these B-cell malignancies support previous observations suggesting that Burkitt's lymphomas are derived through malaria–parasite–antigen driven mechanisms [28] and that VDJ regions from multiple myelomas contain somatic mutations [29]. Our observations on the active immunoglobulin genes of MC116 suggest that undifferentiated B-cell lymphomas have limited somatic mutation of their immunoglobulin genes, which would argue against an antigen driven mechanism of transformation in these tumors. The extent to which both V_H and V_L genes show the same degree of divergence from the nearest germline V-region genes suggests that the mechanisms of somatic mutation that give rise to these changes act in parallel on both heavy and light chain variable region genes.

ACKNOWLEDGMENTS

We thank Robert C. Gallo for his support and encouragement and Linda Anderson for expert editorial assistance.

REFERENCES

1 Huse WD, Shastry L, Iverson SA *et al.* Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. Science 1989;246:1275–81.

2 Kang AS, Barbas CF, Janda KD, Benkovik SJ, Lerner RA. Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. Proc Natl Acad Sci USA 1991;88:4363–6.

3 Barbas CF, Kang AS, Lerner, RA, Benkovik, SJ. Assembly of combinatorial antibody libraries on phage surfaces: The gene III site. Proc Natl Acad Sci USA 1991;88:7978–82.

4 Barbas CF, Lerner RA. Combinatorial immunoglobulin libraries on the surface of phage (Phabs): Rapid selection of antigen specific clones. Methods: A companion to methods in enzymology 1991;2:119–24

5 Persson MAA, Caothien RH, Burton DR. Generation of diverse high-affinity human monoclonal antibodies by repertoire cloning. Proc Natl Acad Sci USA 1991;88:2432–6.

6 Klein E, Klein G, Nadkarni JS, Nadkarni JJ, Wigzell H, Clifford P. Surface IgM-κ specificity on a Burkitt lymphoma cell *in vivo* and in derived cell cultures Cancer Res 1968;28:1300–10.
7 Matsuoka Y, Moore GE, Yagi Y, Pressman D. Production of free light chains of immunoglobulin by a hematopoietic cell line derived from a patient with multiple myeloma. Proc Soc Exp Biol(NY) 1968;125:1246–50.
8 van Boxel JA, Buell DN. IgD on cell membranes of human lymphoid cell lines with multiple immunoglobulin classes. Nature 1974;251:443–4.
9 Magrath IT, Freeman CB, Pizzo P *et al.* Characterization of lymphoma-derived cell lines: Comparison of cell lines positive and negative for Epstein-Barr Virus nuclear antigen J Natl Cancer Inst 1980;64:465–76.
10 Steinitz M, Klein G. EBV-Transformation of Surface IgA positive Human Lymphocytes. J Immunol 1980;125:194–6.
11 Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor University Press 1989.
12 Klobeck H-G, Combriato G, Zachau HG. Immunoglobulin genes of the κ light chain from two human lymphoid cell lines are closely related. Nucleic Acids Research 1984;12:6995–7006.
13 Scott MG, Crimmins DL, McCourt D *et al.* Clonal characterization of the human IgG antibody repertoire to Haemophilus influenzae type b polysaccharide: IV The less frequently expressed V_L are heterologous. J Immunol 1991;147:4007–13.
14 Huber C, Schable KF, Huber E *et al.* The Vκ sequences of the L chains and the repertoire of Vκ gene sequences in the human germ line. Eur J Immunol 1993;23:2868–75.
15 Bentley DL, Rabbitts TH. Human immunoglobulin variable region genes—DNA sequences of two V_κ genes and a pseudogene. Nature 1980;288:730–3.
16 Williams SC, Winter G. Cloning and sequencing of human immunoglobulin V_λ gene segments. Eur J Immunology 1993; 23:1456–61.
17 Combriato G, Klobeck H-G. V_λ and J_λ-C_λ gene segments of the human immunoglobulin λ light chain genes are separated by 14kb and rearrange by a deletion mechanism. Eur J Immunol 1991;21:1513–22.
18 Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. The repertoire of human germline V_H sequences reveals about fifty groups of V_H segments with different hypervariable loops. J Mol Biol 1992;227:776–98.
19 Rechavi G, Bienz B, Ram D *et al.* Organization and evolution of immunoglobulin V_H gene subgroups. Proc Natl Acad Sci USA 1982;79:4405–9.
20 Tomlinson, IM, Cook, GP, Carter, NP *et al.* Human immunoglobulin V_H and D segments on chromosome 15q112 and 16p112. Hum Mol Genet 1994;3:853–60.
21 Keuppers R, Fischer U, Rajewsky K, Gause A. Immunoglobulin heavy and light chain gene sequences of a human CD5 positive immunocytoma and sequences of four novel V_HIII germline genes. Immunol Lett 1992;34:57–62.
22 Guillaume T, Rubinstein DB, Young F *et al.* Individual V_H genes detected with oligonucleotide probes from the complementarity determining regions. J Immunol 1990;145:1934–45.
23 Matthyssens G, Rabbitts TH. Structure and multiplicity of gene for the human immunoglobulin heavy chain variable region. Proc Natl Acad Sci USA 1980;77:6561–5.
24 Marks JD, Tristem, M, Karpas, A, Winter G. Oligonucleotide primers for the polymerase chain reaction amplification of human immunoglobulin variable region genesand design of family specific oligonucleotide probes. Eur J Immunol 1991;21:985–91.
25 de Boer M, Chang S-Y, Eichinger G, Wong HC. Design and analysis of PCR primers for the amplification and cloning of human immunoglobulin Fab fragments. Hum Antibod Hybridomas 1994;5:57–64.
26 Hay BN, Sorge JA, Shopes B. Bacteriophage cloning and *Escherichia coli* expression of a human IgM Fab. Hum Antibod Hybridomas 1992;3:81–5.
27 Zhu D, Hawkins RE, Hamblin TJ, Stevenson FK. Clonal history of a human follicular lymphoma as revealed in the immunoglobulin variable region genes. Br J Haematol 1994;86:505–12.
28 Bakkus MHC, Heirman C, Riet IV, Camp BV, Thielemans K. Evidence that the multiple myeloma Ig heavy chain VDJ genes contain somatic mutations but show no intraclonal variation. Blood 1992;80:2326–35.
29 Magrath, I. The pathogenesis of Burkitt's lymphoma. Adv Cancer Res 1990;55:133–270.

Received 28 April 1995
Accepted in revised form 1 June 1995